COMPOSITIONS AND METHODS FOR THE ANALYSIS OF MUCIN GENE EXPRESSION AND IDENTIFICATION OF DRUGS HAVING THE ABILITY TO INHIBIT MUCIN GENE EXPRESSION

[0001] This invention was made with Government support by Grant Nos. HL35635, ES06230 and ES09701, awarded by the National Institutes of Health. The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Field of the Invention

[0002] The invention relates to compositions and methods for the assessment of mucin gene expression. The invention also relates to compositions and methods for the identification of compounds useful in the treatment of various medical conditions caused by mucin overproduction.

Description of the Related Art

[0003] Mucins are a family of high molecular weight glycoproteins secreted from epithelial cells at many body surfaces, including the eyes, pancreatic ducts, gallbladder, prostate and respiratory, gastrointestinal and reproductive tracts. Mucins are a major component of mucus, and are responsible for the viscoelastic properties of mucus, and serve a role in protecting and lubricating the epithelial surfaces. At least twelve mucin genes have been identified in humans.

[0004] In the airways, mucin proteins form a protective barrier on the airway epithelial cells, and interact with cilia to trap and clear pathogens (*e.g.*, microorganisms), particulate matter, irritants and pollutants (*e.g.*, tobacco smoke and sulphur dioxide). Mucus secretions in the airway are produced from two different secretory cell populations, the surface epithelial goblet cells and the mucous cells in the submucosal glands. At least eight mucin genes are expressed (at the mRNA level) in the upper and lower respiratory tracts. Of these, only the *MUC5AC* and *MUC5B* polypeptides have been conclusively demonstrated to

be major components of human airway secretions (Hovenberg *et al.*, *Biochem. J.*, 318(Pt. 1, Vol. 17):319-324 [1996]; Hovenberg *et al.*, *Glycoconjugate Jour.*, 13(5):839-847 [1996]; Thornton *et al.*, *J. Biol. Chem.*, 272(14):9561-9566 [1997]; and Wickström *et al.*, *Biochem. Jour.*, 334(Pt. 3, Vol. 14):685-693 [1998]). *MUC5B* is also expressed in other tissues, including, for example, pancreas and gall bladder.

Diseases of Mucin Overproduction

[0005] Mucin production is upregulated in response to mucosal irritation. Most notably, bacterial infection of the airway epithelium is often accompanied by mucin overproduction. Some airway diseases are also characterized by mucus hypersecretion. Hypersecretion of mucus can overwhelm the ability of the cilia to function properly, and can result in various pathologies, such as airway mucus plugging and airflow obstruction. Mucus hypersecretion also contributes to chronic infection by shielding bacteria from endogenous and exogenous antibacterial agents. Mucus plugging and bacterial infections create a non-healing injury and can result in chronic influx of inflammatory cells which destroy gas exchange tissue. When severe, these effects result in respiratory function decline, and can be fatal.

[0006] Diseases which are characterized by mucin (and mucus) hypersecretion also frequently demonstrate goblet cell hyperplasia and submucosal gland hypertrophy. Such diseases include, for example, chronic bronchitis, bronchial pneumonia, cystic fibrosis, chronic asthma, emphysema, usual interstitial pneumonitis and other diseases (Basbaum *et al.*, *Am. Rev. Respir. Dis.*, 144(3 Pt 2):S38-41 [1991]; Yanagihara *et al.*, *Am. J. Respir. Cell. Mol. Biol.*, 24(1):66-73 [2001]; Rogers *et al.*, *Eur. Respir. J.*, 7(9):1690-706 [1994]; and Kaliner *et al.*, *American Review of Respiratory Disease* 134(3):612-21[1986]).

MUC5B mRNA and Genomic Structure

[0007] In order to better understand the molecular mechanism of mucin gene expression regulation in normal and disease states, it is necessary to elucidate the genomic structure of the mucin gene. *MUC5B* and three other mucin genes, *MUC6, MUC2,* and *MUC5AC,* have all been mapped to 11p15.5 on a single band of 400 kilobases, and their

order has been determined to be: telomere-*MUC6-MUC2-MUC5AC-MUC5B*-centromere. The *MUC5B* genomic structure (*i.e.*, exon identification, intron/exon boundaries and transcriptional start sites) and cDNA sequence are also partially known, albeit with some discrepancies in the published literature (Pigny *et al.*, *Genomics* 38(3):340-352 [1996]; Desseyn *et al.*, *Jour. Biol. Chem.*, 273(46):30157-30164 [1998]; Desseyn *et al.*, *Jour. Biol. Chem.*, 272(27):16873-16883 [1997]; Desseyn *et al.*, *Jour. Biol. Chem.*, 272(6):3168-3178 [1997]; Offner *et al.*, *Biochem. Biophy. Res. Comm.*, 251(1):350-355 [1998]; and Keates *et al.*, *Biochem. J.*, 324(Pt 1):295-303 [1997]).

[0008] The *MUC5B* gene is large and complex. The *MUC5B* exons and introns encompass approximately 39076 basepairs of genomic sequence, and the gene's cDNA is approximately 17079 basepairs in length. The gene is characterized by an unusually large central exon of 10,713 basepairs and 3,571 amino acids. The central exon contains multiple repeated motifs, including characteristic cysteine-rich subdomains, which are also found in other mucin genes. In addition to the large central exon, there are approximately 30 smaller exons upstream and another approximately 17 exons downstream of the central exon. In total, the *MUC5B* message is predicted to encode a 5683 amino acid polypeptide having a molecular weight of 590 kDa. However, as the mucin proteins are extensively glycosylated, the observed molecular weight is expected to be much greater. Conflicting descriptions of the gene's transcription start sites and identity of the first exon have been reported.

[0009] There exist published reports of the isolation and analysis of limited portions of the *MUC5B* 5' promoter region. Van Seuningen *et al.* (*Biochem. J.*, 348(Pt. 3):675-686 [2000]) describe an analysis of the *MUC5B* promoter region, which encompasses approximately 956 basepairs of genomic nucleotide sequence upstream of the transcription start site. Perrais *et al.* (*J. Biol. Chem.*, 276(18):15386-15396 [2001]) describe an analysis of the *MUC5B* promoter region, which includes approximately 2044 basepairs of genomic nucleotide sequence upstream of the transcription start site. GenBank Accession Number AJ012453 describes approximately 2954 basepairs of *MUC5B* genomic sequence 5' of the transcriptional start site.

[0010] There is a need to identify compounds capable of inhibiting the production of mucin proteins, and specifically, MUC5B protein. There is a need to provide therapies for

reducing mucus (*e.g.*, MUC5B) production in individuals suffering from airway diseases characterized by mucus hypersecretion, such as cystic fibrosis, chronic bronchitis, bronchial pneumonia and asthma. The object of the present invention is to provide novel compositions and methods that find use in the analysis of *MUC5B* gene expression. These compositions incorporate previously unreported *MUC5B* genomic sequences derived from the *MUC5B* gene 5' promoter region, and the methods of the invention use these sequences. These novel compositions further comprise reporter genes in operable combination with the novel *MUC5B* gene 5' promoter regions of the present invention. It is also an object of the present invention to provide methods for drug screening using the novel *MUC5B* promoter reporter constructs to identify compounds having the ability to downregulate *MUC5B* gene expression. The invention also provides transgenic animals suitable for use in screening assays to identify compounds capable of inhibiting mucin production. Compounds thus identified find use in the treatment of diseases characterized by mucin hypersecretion.

SUMMARY OF THE INVENTION

[0011] The present invention provides novel isolated nucleic acid molecules comprising promoter sequences regulating the transcription of the human *MUC5B* gene. These novel sequences are provided in SEQ ID NO: 31 and SEQ ID NO: 32. In a related embodiment, the invention also provides nucleic acid molecules wherein the promoter sequences of SEQ ID NO: 31 or SEQ ID NO: 32 are operably linked to a heterologous gene (*i.e.*, a gene that is not naturally linked to the promoter sequences of SEQ ID NO: 31 or SEQ ID NO: 32).

[0012] In one embodiment, the combination of promoter sequence and heterologous gene reside within a vector. In some embodiments, the heterologous gene contained on the vector is a reporter gene. The heterologous gene can encode various polypeptides, including luciferase, green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT), β-glucuronidase (GUS), secreted alkaline phosphatase (SEAP) and β-galactosidase (β-gal).

[0013] It is intended that host cells harboring the nucleic acid molecules and various vectors of the present invention are also within the scope of the invention. The

nature of the host cell is not particularly limited. In some embodiments, host cells harboring the nucleic acid molecule comprising either promoter sequences of SEQ ID NO: 31 or SEQ ID NO: 32 operably linked to a heterologous gene are provided by the present invention. Furthermore, host cells harboring a vector carrying either of these promoter sequences operably linked to a heterologous gene are also provided by the invention. In related embodiments, host cells harboring a vector carrying either of these promoter sequences operably linked to a reporter gene are provided by the invention. In some embodiments, the host cell is a eukaryotic cell. In other embodiments, the host cell is a cell of human origin. In some preferred embodiments, the host cell is a cell of tracheobronchial epithelial (TBE) origin. When cells are of TBE origin, they may be primary TBE cells or established HBE1 cells. In one embodiment, when the host cells are eukaryotic cells, the host cell can be present in a non-human mammal, in which case the non-human mammal is a transgenic animal. It is intended that transgenic animals comprising the nucleic acid molecules, vectors and host cells of the invention are within the scope of the invention.

[0014] The present invention provides a variety of cell culture conditions and culture methods for the cultivation of the host cells of the invention. In its broadest sense, the invention provides a method for culturing a host cell in a culture medium under conditions allowing the expression of a heterologous gene product that is under the transcriptional control of *MUC5B* promoter sequences SEQ ID NO: 31 or SEQ ID NO: 32. In one embodiment of these cell culture methods, the host cell is of tracheobronchial epithelial (TBE) origin. In other embodiments, the host cell of TBE origin is cultured biphasically in an air-liquid interface. In still other methods for culturing host cells of the invention, the host cell of TBE origin is cultured on a substrate comprising collagen gel. In still other culture methods, the host cells are cultured in the presence of retinoic acid.

[0015] In another embodiment, the present invention provides non-human transgenic mammals comprising eukaryotic host cells harboring the promoter sequences of SEQ ID NO: 31 or SEQ ID NO: 32 operably linked to a heterologous gene.

[0016] The present invention provides a wide variety of methods for the assessment of *MUC5B* promoter activity, and related screening methods to identify compounds having the ability to inhibit human *MUC5B* promoter activity. In one

embodiment, a method for the assessment of *MUC5B* gene promoter activity entails delivering a reporter construct driven by *MUC5B* promoter sequences SEQ ID NO: 31 or SEQ ID NO: 32 operably linked to a reporter gene to a host cell, and assessing the expression of said marker gene product encoded by the reporter gene. In this method, expression of the marker gene product is indicative of *MUC5B* gene promoter activity.

[0017] In a related embodiment, the method above further comprises measuring the quantity of the marker gene product, where the quantity of the marker gene product is proportionate to *MUC5B* gene promoter activity.

[0018] In another embodiment, the present invention provides a method for identifying a compound capable of modulating *MUC5B* gene promoter activity, where the method has the steps of providing a first and a second sample of a host cell, where the host cell harbors a reporter construct driven by a *MUC5B* nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32, operably linked to a reporter gene encoding a marker gene product; contacting the first sample of host cells with a test compound; assessing the expression of the marker gene product in the first and second samples; and identifying the compound as capable of modulating *MUC5B* gene promoter activity if the expression of the marker gene product is significantly different in the first and second samples.

[0019] In a related embodiment of the method above, the quantity of the marker gene product is measured, where the quantity is proportionate to *MUC5B* gene promoter activity. Also in a related embodiment of the method above, the modulation is inhibition.

[0020] The present invention also provides a method for identifying a compound capable of modulating *MUC5B* gene promoter activity. In one embodiment, this method comprises the steps of providing a host cell harboring a reporter construct driven by a *MUC5B* nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32, operably linked to a reporter gene encoding a marker gene product; contacting the host cell with a test compound; measuring the activity of the reporter gene construct; and identifying a compound as capable of modulating *MUC5B* gene promoter activity, if the activity of the reporter gene construct is significantly different from activity measured prior to contact with the test compound. In one embodiment of this method, the modulation is inhibition.

[0021] The present invention provides methods for producing a non-human transgenic animal. In one embodiment, the method comprises the steps of introducing a vector comprising a reporter gene under control of a *MUC5B* promoter sequence comprising a nucleotide sequence of SEQ ID NO: 31 or SEQ ID NO: 32 into an embryonic stem cell of a non-human transgenic animal to produce a transgenic embryonic stem cell; introducing the transgenic embryonic stem cell into a female mouse under conditions such that the mouse delivers progeny of the transgenic embryonic stem cell; and identifying at least one offspring of the progeny comprising the vector.

[0022] In another embodiment of this method, the non-human transgenic animal selectively expresses the reporter gene in a cell of tracheobronchial epithelial (TBE) origin. In another embodiment, the transgenic animal is a mouse.

[0023] The present invention provides methods for screening compounds for the ability to modulate *MUC5B* gene promoter activity. This method comprises the steps of administering a test compound to a non-human transgenic animal produced by the method above, and monitoring *MUC5B* gene promoter activity. In one embodiment of this method, the modulation is inhibition.

[0024] The present invention also provides a method for the specific expression of a nucleic acid of interest in cells of tracheobronchial epithelial (TBE) origin of a mammal, comprising delivering a vector comprising the nucleic acid of interest under control of a *MUC5B* promoter sequence with a sequence of SEQ ID NO: 31 or SEQ ID NO: 32 to the mammal.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] FIGS. 1A-1C show light microscopy images of *in situ* nucleic acid hybridization of human bronchial tissue cross sections from a patient with no obvious airway disease or inflammation. A 48-mer oligonucleotide (SEQ ID NO: 1) corresponding to the antisense sequence of the human *MUC5B* tandem repeats region was used as the *in situ* probe. FIG. 1A shows a section of bronchial tissue after the *in situ* hybridization. Original magnification is 100X. FIG. 1B shows an enlarged picture of the surface epithelium in a region different from FIG. 1A. Original magnification is 400X. FIG. 1C shows an enlarged

picture of the submucosal gland region from FIG. 1A corresponding to the rectangle in that image. Original magnification is 400X.

[0026] FIGS. 2A-2C show light microscopy images of normal and disease airway tissue cross sections following Alcian blue-periodic acid-Schiff (AB-PAS) staining. FIG. 2A shows a normal trachea tissue section following staining. FIG. 2B shows trachea tissue section of a typical interstitial pneumonitis (UIP) patient following staining. FIG. 2C shows a section of bronchiole region tissue from a UIP patient following staining.

[0027] FIGS. 3A-3D show light microscopy images of *in situ* nucleic acid hybridizations of human bronchial tissue cross sections from patients with UIP or emphysema. FIG. 3A shows a section of the trachea tissue of a UIP patient after the *in situ* hybridization. A 48-mer oligonucleotide (SEQ ID NO: 1) corresponding to the antisense sequence of the human *MUC5B* tandem repeats region was used as the *in situ* probe. Original magnification was 100X. FIG. 3B shows a cross section of surface epithelium of the bronchiole region of the UIP paitent's lung. A *MUC5B* oligonucleotide as described in FIG. 3A was used as the *in situ* probe. Original magnification was 400X. FIG. 3C shows an *in situ* hybridization in a human tracheal tissue section derived from a patient with emphysema. A *MUC5B* oligonucleotide as described in FIG. 3A was used as the *in situ* probe. Original magnification was 100X. FIG. 3D shows an *in situ* hybridization in a human tracheal tissue section derived from a patient with emphysema using a *MUC5AC* nucleic acid probe (SEQ ID NO: 2). Original magnification was 100X.

[0028] FIGS. 4A-4B show Northern blot analyses of *MUC5B* message expression in various human cell cultures. The top portions of these blots are probed using a 48 basepair ^{32}P-end labeled nucleic acid probe derived from the repetitive repeat region of the human *MUC5B* gene. FIG. 4A, top panel, shows Northern blot analysis of total RNA isolated from primary explant human tracheobronchial epithelial (TBE) cell cultures. These cultures were maintained under four different culture conditions, which were standard tissue culture dishes (TC), collagen gel coated dishes (CG), Transwell™ chambers (BI), or collagen-gel coated Transwell™ chambers (BICG). Cultures were grown either in the presence (+RA) or absence (-RA) of retinoic acid at a concentration of 30 nM. FIG. 4B, top panel, shows a Northern blot using total RNA isolated from airway cultures and probed for *MUC*5B

message expession. Cells used in the analysis were primary TBE cells, HBE1 cells and BEAS-2B (S clone) cells. The cells used in FIG. 4B were plated using BICG culture conditions contained 30 nM retinoic acid. Following analysis with the *MUC5B* probe, the blots used in FIGS. 4A and 4B were stripped and reprobed with an 18S rRNA cDNA probe as a reference for RNA loading normalization.

[0029] FIGS. 5A and 5B show schematic representations of the Cos-1 cosmid clone and the genomic organization of the amino-terminal and 5' flanking regions of the *MUC5B* gene. FIG. 5A shows the organization of genomic sequences contained on the Cos-1 cosmid clone. The regions corresponding to *MUC5B* and *MUC5AC* coding sequences are shown as filled bars. The 22,773 basepair portion of Cos-1 that was sequenced is indicated. FIG. 5B shows the detailed genomic organization of that part of Cos-1 that was subjected to sequence analysis and which contains *MUC5B* gene exons upstream of the large central exon as well as promoter sequences. Open bars and numbers indicate the exons and the size of these bars are approximately proportional to the relative sizes of the exons. The TATA box, 5' untranslated region (UTR), the initiator ATG, and large central exon are indicated.

[0030] FIG. 6 shows 22,773 basepairs of human *MUC5B* genomic region isolated and sequenced from the Cos-1 genomic cosmid clone. This 22.7 kB encompasses 4169 basepairs of sequence upstream of the transcription start site, the 5'-UT and the 30 exons/introns upstream of the *MUC5B* large central exon.

[0031] FIG. 7 shows a denaturing polyacrylamide gel electrophoresis (PAGE) containing a primer extension analysis of the *MUC5B* transcript. The primer used in the analysis is the Pel1 primer (SEQ ID NO: 7; and TABLE 2). The extension product shown in lane 1 used RNA template isolated from human trachea tissue. The extension product shown in lane 2 used RNA template isolated from human primary tracheobronchial epithelial (TBE) cells. Lanes 3-6 contain a Sanger dideoxynucleotide sequencing ladder in the order GATC, which was produced using a pcDNA3 vector as the nucleic acid template and the Pel1 primer. Radio-labeled dephosphorylated DNA size markers (pBR322/MspI; New England Biolabs, Inc. Beverly, MA) were also run, and whose sizes are indicated on the right.

[0032] FIG.8 shows the nucleotide sequence of the *MUC5B* gene 5'-UTR, adjacent promoter proximal flanking region and the first exon. Only 2007 basepairs of the

sequenced 22,773 basepairs are shown. Various putative DNA motifs are underlined. The transcription start site is indicated by an arrow. The predicted first exon coding region is underlined, and the corresponding predicted signal peptide amino acid sequence is shown using standard letter codes.

[0033] FIG. 9 shows a schematic of the chimeric promoter-reporter gene constructs made using the isolated *MUC5B* gene sequences. The chimeric constructs are termed MUC5B-b1, MUC5B-b2, and MUC5B-il. Each construct contains the luciferase reporter gene and various extents of *MUC5B* promoter-proximal sequences.

[0034] FIG. 10 shows the *MUC5B* genomic nucleotide sequence encompassing positions -1098 through +7 that were subcloned into the MUC5B-b1 luciferase reporter construct.

[0035] FIG. 11 shows the *MUC5B* genomic nucleotide sequence encompassing positions -4169 through +7 that were subcloned into the MUC5B-b2 luciferase reporter construct.

[0036] FIG. 12 shows the *MUC5B* genomic nucleotide sequence encompassing positions -13 through +2738 that were subcloned into the MUC5B-il luciferase reporter construct.

[0037] FIG. 13 shows the results of a transfection assay using the chimeric *MUC5B* luciferase reporter constructs shown in FIG. 9 and primary TBE cells. The TBE cells were also co-transfected with a β-galactosidase expression vector, and luciferase activity was normalized against β-galactosidase activity to take into account transfection efficiency variability. Relative activities of each of the reporter constructs following transfection in the TBE cells is shown, and activity is expressed as units of luciferase activity per unit of β-gal activity (units/beta-gal).

[0038] FIG. 14 shows the results of a transfection assay using the MUC5B-b2 luciferase reporter construct shown in FIG. 9 and three different cell types. These were primary TBE cells (unfilled bars), HBE1 cells (striped bars) and BEAS-2B (S clone) cells (black bars). The cells were also co-transfected with a β-galactosidase expression vector, and luciferase activity was normalized against β-galactosidase activity to take into account transfection efficiency variability, and activity is expressed as units of luciferase activity per

unit of β-gal activity (units/beta-gal). Transfections were done in triplicate, and the mean results of two independent experiments are shown.

[0039] Fig. 15 shows the results of a transfection study using the MUC5B-b2 luciferase reporter construct shown in FIG. 9 and primary human TBE cells. The TBE cells were maintained in two different culture conditions, which were standard tissue culture dishes (TC) or collagen gel-coated Transwell™ chambers (BICG). Activity of the MUC5B-b2 reporter construct was observed in the cultures maintained in the presence (+RA) or absence (-RA) of retinoic acid. The luciferase reporter gene activity in each transfected culture was normalized to the activity of a cotransfected β-galactosidase expression vector. Results are expressed as "fold increase" of luciferase activity, comparing RA-treated and RA-untreated cultures. The activity of the MUC5B-b2 reporter in RA-untreated culture in the TC conditions was normalized to 1. Transfections were done in triplicate, and the mean results of two independent experiments are shown.

[0040] FIG. 16 shows the results of an analysis of MUC5B-b2 luciferase reporter activity in the context of stably integrated cells derived from transgenic animals. Transgenic mice carrying the MUC5B 4,169/+7 promoter luciferase reporter were used to isolate and culture primary TBE cells. The TBE cultures were maintained in three different conditions, which were with interleukin-6, with interleukin-12 or without any interleukin (control). Cells were harvested and extracts prepared. Luciferase activity was determined for each culture, normalized for total protein in the samples, and graphed.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0041] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Indeed, the present invention is in no way limited to the methods and materials described. For purposes of the present invention, the following terms are defined below.

[0042] The terms "nucleic acid," "nucleic acid sequence," "nucleotide sequence," "oligonucleotide," "polynucleotide" or "nucleic acid molecule" as used herein refer to an oligonucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which can be single- or double-stranded, and represent the sense or antisense strand. The terms nucleic acid, polynucleotide and nucleotide also specifically include nucleic acids composed of bases other than the five biologically occurring bases (*i.e.*, adenine, guanine, thymine, cytosine and uracil).

[0043] As used herein, the term "oligonucleotide," refers to a short length of single-stranded polynucleotide chain. Oligonucleotides are typically less than 100 nucleotides long (e.g., between 15 and 50), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer." Oligonucleotides can form secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides.

[0044] As used herein, "recombinant nucleic acid," "recombinant gene" "recombinant DNA molecule" or similar terms indicate that the nucleotide sequence or arrangement of its parts is not a native configuration, and has been manipulated by molecular biological techniques. The term implies that the DNA molecule is comprised of segments of DNA that have been artificially joined together. Protocols and reagents to produce recombinant nucleic acids are common and routine in the art *(See e.g.,* Maniatis *et al.*(eds.), *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory Press, NY, [1982]; Sambrook *et al.* (eds.), *Molecular Cloning: A Laboratory Manual*, Second Edition, Volumes 1-3, Cold Spring Harbor Laboratory Press, NY, [1989]; and Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology, Vol.* 1-4, John Wiley & Sons, Inc., New York [1994]).

[0045] As used herein, the term "probe" refers to an oligonucleotide (*i.e.,* a sequence of nucleotides), which is often produced from nucleic acid isolated from cells (typically a recombinant nucleic acid), produced synthetically or *in vitro,* which is capable of hybridizing to a nucleic acid of interest. Probes are useful in the detection, identification and isolation of particular gene or mRNA sequences. It is contemplated that any probe used in

the present invention is capable of being labeled with any "reporter molecule," so that the probe is detectable. Detection systems include, but are not limited to, the detection of enzymatic activity, fluorescence, radioactivity, and luminescence. In addition, a detection system may also comprise a specific antibody. It is not intended that the present invention be limited to any particular probe, label or detection system.

[0046] The terms "peptide," "polypeptide" and "protein" all refer to a primary sequence of amino acids that are joined by covalent "peptide linkages." In general, a peptide consists of a few amino acids, typically from 2-25 amino acids, and is shorter than a protein. "Polypeptides" encompass both peptides or proteins. As used herein, a recited "amino acid sequence" refers to an amino acid sequence of a naturally occurring protein molecule, a protein produced by recombinant molecular genetic techniques, or a synthetic or naturally occurring peptide, and may refer to a portion of a larger "peptide," "polypeptide" or "protein," and is not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

[0047] The terms "exogenous" and "heterologous" are sometimes used interchangeably with "recombinant." An "exogenous nucleic acid," "exogenous gene" and "exogenous protein" indicate a nucleic acid, gene or protein, respectively, that has come from a source other than its native source, and has been artificially supplied to the biological system. In contrast, the terms "endogenous protein," "native protein," "endogenous gene," and "native gene" refer to a protein or gene that is native to the biological system, species or chromosome under study. A "native" or "endogenous" gene is a gene that does not contain nucleic acid elements encoded by sources other than the chromosome on which it is normally found in nature. An endogenous gene or transcript is encoded by its natural chromosomal locus, and not artificially supplied to the cell.

[0048] The term "isolated" when used in relation to a nucleic acid, as in "an isolated nucleic acid," "an isolated oligonucleotide," "isolated polynucleotide" or "isolated nucleotide sequence," refers to a nucleic acid that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is present in a form or setting that is different from the form or setting of that nucleic acid found in nature. In contrast, non-isolated nucleic acids are found in the state in

which they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell in a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid encoding a given polypeptide includes, by way of example, such nucleic acid in cells ordinarily expressing the given protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. This isolated nucleic acid, oligonucleotide, or polynucleotide is either single-stranded or double-stranded. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (*i.e.*, the oligonucleotide or polynucleotide is single-stranded). In other embodiments, the oligonucleotide or polynucleotide contains both the sense and anti-sense strands (*i.e.*, the oligonucleotide or polynucleotide is double-stranded).

[0049] As used herein, the term "purified" or "to purify" refers to the removal of at least one contaminant from a sample. As used herein, the term "substantially purified" refers to molecules, either nucleic acids or amino acid sequences, that are removed from their natural environment, "isolated" or "separated," and are largely free from other components with which they are naturally associated. An "isolated nucleic acid" or "isolated polypeptide" are therefore a substantially purified nucleic acid or substantially purified polypeptide.

[0050] Nucleic acid molecules (*e.g.*, DNA or RNA) are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides or polynucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotides or polynucleotide, referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide or polynucleotide, also can be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete

elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements that direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, in some embodiments, enhancer elements exert their effect even when located 3' of the promoter element or the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

[0051] The term "gene" refers to a nucleic acid (*e.g.*, DNA) sequence comprised of parts, that when appropriately combined in either a native or recombinant manner, provide some product or function. In some embodiments, genes comprise coding sequences necessary for the production of a polypeptide, while in other embodiments, the genes do not comprise coding sequences necessary for the production of a polypeptide. Examples of genes that do not encode polypeptide sequences include ribosomal RNA genes (rRNA) and transfer RNA (tRNA) genes. In preferred embodiments, genes encode a polypeptide or any portion of a polypeptide within the gene's "coding region" or "open reading frame." In some embodiments, the polypeptide produced by the open reading frame of a gene displays at least one functional activity (*e.g.*, enzymatic activity, ligand binding, signal transduction, etc.), while in other embodiments, it does not.

[0052] In addition to the coding region of the nucleic acid, the term "gene" also encompasses the transcribed nucleotide sequences of the full-length mRNA adjacent to the 5' and 3' ends of the coding region. These noncoding regions are variable in size, and typically extend for distances up to or exceeding 1 kb on both the 5' and 3' ends of the coding region. The sequences that are located 5' and 3' of the coding region and are contained on the mRNA are referred to as 5' and 3' untranslated sequences (5' UT and 3' UT). Both the 5' and 3' UT may serve regulatory roles, including translation initiation, post-transcriptional cleavage and polyadenylation. The term "gene" encompasses mRNA, cDNA and genomic forms of a gene.

[0053] In some embodiments, the genomic form or genomic clone of a gene contains the sequences of the transcribed mRNA, as well as other non-transcribed sequences which lie outside of the mRNA. The regulatory regions which lie outside the mRNA

transcription unit are sometimes called "5' or 3' flanking sequences." A functional genomic form of a gene must contain regulatory elements necessary for the regulation of transcription. The term "promoter/enhancer region" is usually used to describe this DNA region, typically but not necessarily 5' of the site of transcription initiation, sufficient to confer appropriate transcriptional regulation. Used alone, the term "promoter" is sometimes used synonymously with "promoter/enhancer." In some embodiments, the promoter is constitutively active, or while in alternative embodiments, the promoter is conditionally active (*i.e.,* where transcription is initiated only under certain physiological conditions or in the presence of certain drugs). In some embodiments, the 3' flanking region contains additional sequences which regulate transcription, especially the termination of transcription. "Introns" or "intervening regions" or "intervening sequences" are segments of a gene which are contained in the primary transcript (*i.e.,* hetero-nuclear RNA, or hnRNA), but are spliced out to yield the processed mRNA form. In some embodiments, introns contain transcriptional regulatory elements such as enhancers. The mRNA produced from the genomic copy of a gene is translated in the presence of ribosomes to yield the primary amino acid sequence of the polypeptide.

[0054] As used herein, the term "regulatory element" refers to a genetic element which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element that enables the initiation of transcription of an operably linked coding region. Other regulatory elements are splicing signals, polyadenylation signals, termination signals, etc.

[0055] Transcriptional control signals in eukaryotes comprise "promoter" and "enhancer" elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis *et al., Science* 236:1237 [1987]). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect and mammalian cells, as well as viruses. Analogous control elements *(i.e.,* promoters and enhancers) are also found in prokaryotes. The selection of a particular promoter and enhancer to be operably linked in a recombinant gene depends on what cell type is to be used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are

functional only in a limited subset of cell types (for review see, Voss *et al., Trends Biochem. Sci.,* 11:287 [1986] and Maniatis *et al., Science* 236:1237 [1987]). For example, the SV40 early gene enhancer is very active in a wide variety of mammalian cell types (Dijkema *et al., EMBO* J, 4:761 -22- [1985]). Two other examples of promoter/enhancer elements active in a broad range of mammalian cell types are those from the human elongation factor t a gene (Uctsuki *et al., J. Biol. Chem.,* 264:5791 [1989]; Kim *et al., Gene* 91:217 [1990]; Mizushima and Nagata, Nuc. *Acids. Res.,* 18:5322 [1990]), the long terminal repeats of the Rous sarcoma virus (Gorman *et al., Proc. Natl. Acad. Sci. USA* 79:6777 [1982]), and human cytomegalovirus (Boshart *et al., Cell* 41:521 [1985]). Some promoter elements serve to direct gene expression in a tissue-specific manner.

[0056] As used herein, the term "promoter/enhancer" denotes a segment of DNA which contains sequences capable of providing both promoter and enhancer functions *(i. e.,* the functions provided by a promoter element and an enhancer element). For example, the long terminal repeats of retroviruses contain both promoter and enhancer functions. In some embodiments, the promoter/enhancer is "endogenous," while in other embodiments, the promoter/enhancer is "exogenous," or "heterologous." An "endogenous" promoter/enhancer is one which is naturally linked with a given gene in the genome. An "exogenous" or "heterologous" promoter/enhancer is one placed in juxtaposition to a gene by means of genetic manipulation (*i.e.,* molecular biological techniques such as cloning and recombination) such that transcription of the gene is controlled by the linked promoter/enhancer.

[0057] The terms "in operable combination," "in operable order," "operably linked" and similar phrases when used in reference to nucleic acids herein are used to refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

[0058] As used herein, the terms "an oligonucleotide having a nucleotide sequence encoding a gene," "polynucleotide having a nucleotide sequence encoding a gene," and similar phrases are meant to indicate a nucleic acid sequence comprising the coding

region of a gene (*i.e.,* the nucleic acid sequence which encodes a gene product). In some embodiments, the coding region is present in a cDNA, while in other embodiments, the coding region is present in genomic DNA or RNA form. When present in a DNA form, the oligonucleotide, polynucleotide or nucleic acid is either single-stranded *(i.e.,* the sense strand or the antisense strand) or double-stranded. In some embodiments, suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, etc. are placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention contains endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, etc. or a combination of both endogenous and exogenous control elements.

[0059] As used herein, the terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" and similar phrases refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid encoding a particular polypeptide. The order of the deoxynbonucleotides determines the order of the amino acids in the polypeptide chain. The DNA sequence thus codes for the amino acid sequence.

[0060] As used herein, the term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (*e.g.,* mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (*i.e.,* via the enzymatic action of an RNA polymerase), and for protein encoding genes, into protein through "translation" of the mRNA. Gene expression regulation often occurs at many stages. "Up-regulation" or "activation" refers to regulation that increases the production of gene expression products *(i.e.,* RNA or protein), while "down-regulation" or "repression" refers to regulation that decreases mRNA or protein production. Molecules (*e.g.,* transcription factors) that are involved in up-regulation or down-regulation are often called "activators" and "repressors," respectively.

[0061] As used herein, the terms "reporter gene" or "reporter" refer to a gene and/or gene product that can be readily detected in a biological system. The choice of the most suitable reporter gene to use for a particular application depends on the intended use, and other variables known to one familiar with the art. Many reporter genes are known in the

art. Each reporter gene has a particular assay for the detection of that reporter. Some detection assays are enzymatic assays, while other assays can be immunological in nature (*e.g.*, ELISA or immunohistochemical analysis).

[0062] As used herein, the term "vector" is used in reference to nucleic acid molecules that cam be used to transfer DNA segment(s) from one cell to another. The terms "vehicle" or "construct" or "plasmid" are sometimes used interchangeably with "vector." In some embodiments, a vector "backbone" comprises those parts of the vector which mediate its maintenance and enable its intended use (*e.g.*, the vector backbone contains sequences necessary for replication, genes imparting drug or antibiotic resistance, a multiple cloning site, and possibly operably linked promoter/enhancer elements which enable the expression of a cloned nucleic acid). The cloned nucleic acid (*e.g.*, such as a cDNA coding sequence, or an amplified PCR product) is inserted into the vector backbone using common molecular biology techniques. Vectors are often derived from plasmids, bacteriophages, or plant or animal viruses. A "cloning vector" or "shuttle vector" or "subcloning vector" contain operably linked parts which facilitate subcloning steps (*e.g.*, a multiple cloning site containing multiple restriction endonuclease sites). A "recombinant vector" indicates that the nucleotide sequence or arrangement of its parts is not a native configuration, and has been manipulated by molecular biological techniques. The term implies that the vector is comprised of segments of DNA that have been artificially joined. A "reporter construct" is a vector encoding a suitable "reporter" gene. The transcription of the reporter gene is typically regulated by heterologous promoter sequences.

[0063] The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and operably linked nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism (*e.g.*, a bacterial expression vector, a yeast expression vector or a mammalian expression vector). Nucleic acid sequences necessary for expression in prokaryotes typically include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells utilize promoters, enhancers, and termination and polyadenylation signals and other sequences which are generally different from those used by prokaryotes.

[0064] The term "transfection" as used herein refers to the introduction of foreign DNA into cells. Transfection can be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, recombinant retroviral infection, and biolistics. Mammalian cell transfection techniques are common in the art, and are described in many sources (*See, e.g.,* Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology,* Chapter 9, John Wiley & Sons, Inc., New York [1994]).

[0065] The term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell which contains stably integrated foreign DNA within its own genomic DNA. A cell that that has been stably transfected transmits the transfected and integrated DNA to all subsequent cell generations, most typically in the presence of a selectable marker.

[0066] The term "transient transfection" or "transiently transfected" refers to the introduction of foreign DNA into a cell where the foreign DNA fails to integrate into the genome of the transfected cell. The foreign DNA persists in the nucleus of the transfected cell for several days. During this time the foreign DNA is subject to the regulatory controls that govern the expression of endogenous genes in the chromosomes. The term "transient transfectant" refers to cells which have taken up foreign DNA but have failed to integrate this DNA.

[0067] The term "calcium phosphate co-precipitation" refers to a technique for the introduction of nucleic acids into a eukaryotic cell, and most typically mammalian cells. The uptake of nucleic acids by cells is enhanced when the nucleic acid is presented as a calcium phosphate-nucleic acid co-precipitate. Various modifications of the original technique of Graham and van der Eb (Graham and van der Eb, Virol., 52:456 [1973]) are known in which the conditions for the transfection of a particular cell type has been optimized. The art is well aware of these various methods.

[0068] The term "transformation" has various meanings, depending on its usage. In one sense, the term "transformation" is used to describe the process of introduction of

foreign DNA into prokaryotic cells (*i.e.*, bacterial cells), and most frequently *E. coli* strains. Bacterial cell transformation can be accomplished by a variety of means well known in the art, including the preparation of "competent" bacteria by the use of calcium chloride, magnesium chloride or rubidium chloride, and electroporation. When a plasmid is used as the transformation vector, the plasmid typically contains a gene conferring drug resistance, such as the genes encoding ampicillin, tetracycline or kanamycin resistance. Bacterial transformation techniques are common in the art, and are described in many sources *(e.g.,* Cohen *et al., Proc. Natl. Acad. Sci. USA* 69: 2110-2114 [1972]; Hanahan, J. *Mol. Biol.,* 166:557-580 [1983]; Sambrook *et al.* (eds.), *Molecular Cloning: A Laboratory Manual,* Second Edition, Volumes 1-3, Cold Spring Harbor Laboratory Press, NY, [1989]; Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology,* Vol. 1-4, John Wiley & Sons, Inc., New York [1994]).

[0069] "Transformation" also describes the physiological process by which a normal eukaryotic cell acquires the phenotypic properties of a malignant cell. Such properties include, but are not limited to the ability to grow in soft agar, the ability to grow in nutrient poor conditions, rapid proliferation, and the loss of contact inhibition. A eukaryotic cell which is "transformed" displays the properties of malignant cells. In some embodiments, eukaryotic cells acquire their transformed phenotype *in vivo,* while in other embodiments, the cells are artificially transformed in culture.

[0070] As used herein, the term "established" or "established culture" is a cell culture, most typically a mammalian cell culture, that has acquired the ability to grow indefinately in culture (in contrast to a primary cell culture). An established cell culture may or may not display traits of transformed cells. Mammalian cells can be established artificially, *e.g.*, by the stable forced expression of the SV-40 large T-antigen.

[0071] As used herein, the term "selectable marker" refers to the use of a gene that encodes an enzymatic activity that confers the ability to grow in medium lacking what would otherwise be an essential nutrient *(e.g.,* the HIS3 gene in yeast cells); in addition, in some embodiments, a selectable marker confers resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed. Furthermore, some selectable markers are "dominant." Dominant selectable markers encode an enzymatic activity that is detectable in

any suitable eukaryotic cell line. Examples of dominant selectable markers include the bacterial aminoglycoside 3' phosphotransferase gene (*i.e.,* the *neo* gene) that confers resistance to the drug G-418 in mammalian cells, as well as the bacterial hygromycin G phosphotransferase *(hyg)* gene that confers resistance to the antibiotic hygromycin, and the bacterial xanthine-guanine phosphoribosyl transferase gene (*i.e.,* the *gpt* gene) that confers the ability to grow in the presence of mycophenolic acid. The use of non-dominant selectable markers must be in conjunction with a cell line that lacks the relevant enzyme activity. Examples of non-dominant selectable markers include the thymidine kinase *(tk)* gene (used in conjunction with *tk-* cell lines), the CAD gene (used in conjunction with CAD-deficient cells) and the mammalian hypoxanthine-guanine phosphoribosyl transferase *(hprt)* gene (used in conjunction with *hprt - cell* lines). A review of the use of selectable markers in mammalian cell lines is provided in Sambrook *et al., Molecular Cloning: A Laboratory Manual,* Second Edition, Cold Spring Harbor Laboratory Press, New York (1989), at pp.16.9-16.15.

[0072] As used herein, the term "cell culture" refers to any *in vitro* culture of cells. Included within this term are continuous cell lines *(e.g.,* with an immortal phenotype), primary cell cultures, finite cell lines (*e.g.,* non-transformed cells), and any other cell population maintained *in vitro*.

[0073] As used herein, the term "eukaryote" refers to organisms distinguishable from "prokaryotes." It is intended that the term encompass all organisms with cells that exhibit the usual characteristics of eukaryotes such as the presence of a true nucleus bounded by a nuclear membrane, within which lie the chromosomes, the presence of membrane-bound organelles, and other characteristics commonly observed in eukaryotic organisms. Thus, the term includes, but is not limited to such organisms as fungi, protozoa, and animals (*e.g.,* humans).

[0074] As used herein, the terms "host," "expression host," and "transformant" refer to organisms and/or cells which harbor an exogenous DNA sequence (*e.g.,* via transfection), an expression vector or vehicle, as well as organisms and/or cells that are suitable for use in expressing a recombinant gene or protein. It is not intended that the present invention be limited to any particular type of cell or organism. Indeed, it is

contemplated that any suitable organism and/or cell will find use in the present invention as a host.

[0075] As used herein, the term "host cell" refers to any cell capable of harboring an exogenous nucleic acid or gene product. In some embodiments, the host cell also transcribes and/or translates and expresses a gene contained on the exogenous nucleic acid. It is intended that the exogenous nucleic acid be obtained from any suitable source. In some embodiments, it is produced synthetically, while in other embodiments, it is produced by another cell or organism. In addition, in some embodiments, the exogenous nucleic acid is subjected to replication, while in other embodiments, it is not.

[0076] As used herein, the term *"in vitro"* refers to an artificial enviromnent and to processes or reactions that occur within an artificial environment. The term "*in vivo*" refers to the natural environment (*e.g.,* in an animal or in a cell) and to processes or reactions that occur within a natural environment. The definition of an *in vitro* versus *in vivo* system is particular for the system under study.

[0077] The term "mammal" or "mammalian species" refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, cats, cattle, horses, sheep, pigs, goats, rabbits, as well as rodents such as mice and rats, etc. Preferably, the mammal is human.

[0078] As used herein, the term "inhibit" refers to the act of diminishing, suppressing, alleviating, preventing, reducing or eliminating. For example, in some embodiments, a compound that inhibits a gene promoter activity results in elimination or reduced transcription of that gene. The term "inhibit" applies equally to both *in vitro* and *in* vivo systems.

[0079] As used herein, the term "chimeric" molecule (*e.g.*, a chimeric plasmid construct or chimeric gene or chimeric protein) refers to a molecule that comprises various elements that are not in a combination normally found in nature. For example, a luciferase reporter open reading frame under the transcriptional control of a MUC5B promoter element can be considered a chimeric gene.

[0080] As used herein, the terms, "primary," "primary culture" or "primary explant" or the like refer to a cell culture, typically a mammalian cell culture, where the cells

in the culture are of low passage number (have not been maintained in culture for an extended period of time following their isolation from an organism) and where the cells are not immortal (*i.e.*, not "established"). In one embodiment, a primary culture is derived from a tissue sample from a human subject.

[0081] The term "cell type specific" as it applies to a gene promoter refers to a promoter that imparts preferential transcriptional activity (*i.e.*, "preferential expression" or "selective expression") onto a downstream nucleic acid in the context of one or a subset of specific cell type(s) relative to another cell type. Preferably, cell specific expression means selective expression of a nucleic acid in one specific tissue, as compared to no significant (or detectable) expression of the same nucleic acid in a different cell type. Cell-type specificity of a promoter can be evaluated in a variety of ways and in various *in vitro* and *in vivo* model systems, as known to one familiar with the art. In one embodiment, the cell type specificity of a promoter is evaluated, for example, by operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into cultured cells (either stably or transiently), and detecting the expression of the reporter gene in various types of cultured cells (*i.e.*, cultured cells of different origins). Selectivity need not be absolute. The detection of a greater level of expression of the reporter gene in one cell type (or a subset of cell types) relative to the level of expression of the reporter gene in other cell type(s) shows that the promoter is specific for the cell type(s) in which greater levels of expression are detected. A single tissue can comprise multiple cell types. The cell types being compared can come from different tissues, or be derived from the same tissue.

[0082] Alternatively, in another embodiment, the cell type specificity of a promoter is evaluated by constructing a suitable reporter construct and introducing the reporter construct into the cells of an animal. The construct can be either stably delivered (in which case the reporter is integrated into the animal genome) or transiently delivered to all cells or a subset of the cells of an animal to form a transgenic animal. The expression of the reporter gene in the cells of that animal is then assessed. The detection of a greater level of expression of the reporter gene in one (or more) cell type relative to the level of expression of the reporter gene in other cell type(s) shows that the promoter is specific for the cell type(s) in which greater levels of expression are detected. Selectivity need not be absolute.

[0083] Preferably, cell type specific expression means selective expression of a nucleic acid in a specific type of cell compared to no significant expression of the same nucleic acid in other types of cells within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting preferential (including selective) expression of a nucleic acid in a region within a single tissue. It is clear from this definition that cell type specificity need not be absolute.

[0084] The term "tissue specific" as it applies to a gene promoter refers to a promoter that imparts preferential transcriptional activity (*i.e.*, preferential expression) onto a downstream nucleic acid in the context of one or a subset of specific tissue type(s) relative to another tissue type. Tissue specificity of a promoter is a function of the cell type specificity of that promoter, where the promoter is more active in the cells of one tissue relative to the cells of a different tissue. A single tissue can comprise multiple cell types. A gene promoter need not be active in every cell type within a given tissue for the promoter to be considered tissue specific. Preferably, tissue specific expression means selective expression of a nucleic acid in one specific tissue, as compared to no significant (or detectable) expression of the same nucleic acid in a different tissue. Selectivity need not be absolute. Tissue specificity of a promoter can be evaluated in a variety of ways and in various *in vitro* and *in vivo* model systems, as known in the art. The detection of a greater level of expression of the reporter gene in one (or more) cell type relative to the level of expression of the reporter gene in other cell type(s) shows that the promoter is specific for the tissues in which greater levels of expression are detected.

[0085] The cell type specificity or tissue specificity of a promoter can be assessed using methods other than reporter constructs, as known in the art. For example, the specificity of a promoter within a cell type, and more commonly within a tissue, can be assessed using *in situ* hybridization techniques with nucleic acid probes, as known in the art. Also, the specificity of a promoter within a tissue can be assessed using immunohistochemical staining. Briefly, when using immunohistochemistry, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the nucleic acid whose expression is controlled by the promoter. A labeled (*e.g.*, peroxidase conjugated) secondary antibody

which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding is visualized and observed microscopically (*e.g.*, by colorimetric visualization of peroxidase activity, and/or by using an avidin/biotin labeling system).

[0086] The terms "selective expression", "selectively express" and grammatical equivalents thereof refer to a comparison of relative levels of expression in two or more regions of interest. For example, "selective expression" when used in connection with tissues refers to a substantially greater level of expression of a gene of interest in a particular tissue, or to a substantially greater number of cells which express the gene within that tissue, as compared, respectively, to the level of expression of, and the number of cells expressing, the same gene in another tissue (*i.e.*, selectivity need not be absolute). Selective expression does not require, although it may include, expression of a gene of interest in a particular tissue and a total absence of expression of the same gene in another tissue. Similarly, "selective expression" as used herein in reference to cell types refers to a substantially greater level of expression of, or a substantially greater number of cells which express, a gene of interest in a particular cell type, when compared, respectively, to the expression levels of the gene and to the number of cells expressing the gene in another cell type.

[0087] The term "promoter activity" when made in reference to a nucleic acid sequence refers to the ability of the nucleic acid sequence to initiate transcription of a downstream deoxyribonucleic acid (DNA) sequence into a ribonucleic acid (*i.e.*, RNA) sequence (*e.g.*, messenger-RNA, transfer-RNA or ribhosomal-RNA).

[0088] The term "sample" as used herein is used in its broadest sense. A "sample" is typically of biological origin, where "sample" refers to any type of material obtained from animals or plants *(e.g.,* any fluid or tissue), cultured cells or tissues, cultures of microorganisms (prokaryotic or eukaryotic), and any fraction or products produced from a living (or once living) culture or cells. A sample can be a cell extract (*i.e.,* a cell lysate), and can be purified or unpurified. An "experimental sample" is a sample where the presence, concentration and/or activity of some molecule of interest is unknown. A "control sample" is a sample where the presence, concentration and/or activity of some molecule of interest is known.

[0089] As used herein, the term "transgene" refers to a nucleic acid sequence which is partly or entirely heterologous, *i.e.*, foreign to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (*e.g.*, it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can be operably linked to one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid. A transgene can also comprise a "reporter gene," which facilitates visualization or quantitation of expression of the transgene.

[0090] Accordingly, the term "transgene construct" refers to a nucleic acid that includes a transgene, and (optionally) such other nucleic acid sequences as transcriptionally regulatory sequence, polyadenylation sites, replication origins, marker genes, etc., which may be useful in the general manipulation of the transgene for insertion in the genome of a host organism.

[0091] The term "transgenic" is used herein as an adjective to describe the property, for example, of an animal or a construct, of harboring a transgene. For instance, as used herein, a "transgenic organism" is any animal, preferably a non-human mammal, in which one or more of the cells of the animal contain heterologous nucleic acid introduced by way of human intervention, such as by trangenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the transgenic animals described herein, the transgene is in the form of a reporte gene, the transcription of which is driven by MUC5B promoter sequences (*e.g.*, SEQ ID NOs: 31 or 32). The terms "founder line" and "founder animal" refer to those animals that are the mature product of the embryos to which the

transgene was added, *i.e.*, those animals that grew from the embryos into which DNA was inserted, and that were implanted into one or more surrogate hosts.

[0092] The terms "progeny" and "progeny of the transgenic animal" refer to any and all offspring of every generation subsequent to the originally transformed mammals. The term "non-human mammal" refers to all members of the class Mammalia except humans. "Mammal" refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as mouse, rat, rabbit, pig, sheep, goat, cattle and higher primates.

Description of the Preferred Embodiments

[0093] In its broadest aspect, the present invention relates to compositions and methods for the analysis of mucin gene expression. The present invention provides the genomic 5' regulatory domain of the human mucin-5B (*MUC5B*) gene. This regulatory domain is used to construct various reporter constructs which find use in drug screening. It is contemplated that *MUC5B* reporter constructs can be used to identify compounds which downregulate (*i.e.*, inhibit) *MUC5B* gene expression. Compounds that are able to downregulate *MUC5B* production find use in the treatment of diseases characterized by mucin hypersecretion and airway plugging.

I. *MUC5B* Overexpression is Observed in Diseased Airway Tissues

[0094] In the present study, *MUC5B* expression was analyzed in normal and diseased airway tissues using *in situ* hybridization techniques, as described in EXAMPLE 2 and FIGS. 1-3. These experiments demonstrated that *MUC*5B message is present in non-diseased tissue, and is predominantly expressed in the submucosal gland cells of tracheobronchial airway tissue (FIG. 1C). However, in airway tissues from patients demonstrating emphysema and ususal interstitial pneumonitis (UIP), there is a general elevated expression of *MUC5B* in the submucosal gland cells, and in addition, *MUC*5B message expression is also present in the surface goblet cell population in diseased lung tissues (*see*, FIGS. 3A-3C). These observations are in agreement with previous reports that suggested that the *MUC5B* gene product was one of the major components in mucus obtained

from asthma (Sheehan *et al.*, *Biochemical Journal* 338(Pt 2)(7):507-513 [1999]) and cystic fibrosis patients (Davies *et al.*, *Biochemical Journal* 344 Pt 2(4697):321-330 [1999]). In contrast to *MUC5B* gene expression, the expression of *MUC5AC* message is restricted to the airway surface epithelium in normal and diseased airway tissues, and does not show elevated expression in disease states. These results suggest a significant positive correlation between elevated *MUC5B* gene expression and the presence of pathogenesis in airway diseases. Such an association was not seen for the expression of *MUC5AC* message (*see*, FIG. 3D).

II. Isolation of *MUC5B* Genomic Sequences

[0095] For the purpose of studying *MUC5B* transcriptional regulation and genomic structure, genomic DNA encompassing the *MUC5B* transcriptional start site was isolated. To isolate genomic DNA clones containing *MUC5B* nucleotide sequence, an initial low-stingency hybridization strategy using a *MUC2* amino-terminal and promoter proximal region nucleic acid probe was used to screen a Clontech human genomic library (the *MUC2* and *MUC5B* genes contain strong homology in their promoter and amino-terminal domains). This initial screening of 10^6 cosmid clones identified eight (8) candidate clones, which were then subjected to a secondary screening using *MUC5AC* cDNA sequences as a Southern blot probe under high stringency conditions. This secondary screen of the initial eight positive clones yielded only a single positive cosmid clone, which was termed Cos-1. The detailed methodology and reaction conditions used in this isolation are provided in EXAMPLE 4.

[0096] This clone was sequenced, and it was found that one end of the clone contained the 5' half of the *MUC5B* coding region, while the opposite end contained coding sequence from the 3' end of the adjacent *MUC5AC* gene. Thus, based on the known gene order on 11p15.5 of *cen-MUC5AC-MUC5B-tel*, it was concluded that the Cos-1 clone must contain the nucleotide sequence corresponding to the 5' promoter region of *MUC5B*.

[0097] The total size of the genomic insert on the Cos-1 clone was estimated to be approximately 44 kB, as determined by restriction mapping (*see*, EXAMPLE 4). Of this 44 kB sequence, the 5' half of the clone accounting for 22,773 basepairs, was fully sequenced. This 22.7 kB encompassed 4169 basepairs upstream of the transcription start site, the 5'-untranslated (5'-UT) region, and the first 30 N-terminal *MUC5B* exons (*i.e.*, all exons/intons

upstream of the large central exon). This sequence was submitted to GenBank (GenBank Accession No. AF107890; and *see*, FIG. 6 and SEQ ID NO: 6). A schematic representation of the Cos-1 clone and genomic organization of the *MUC5B* gene upstream of the large central exon is shown in FIG. 5.

[0098] Another depiction of part of the 22.7 kB sequence proximal to the transcription start site showing predicted landmarks of the gene is shown in FIG. 8. This Figure shows the predicted *MUC5B* transcription start site, a TATA box 30 nucleotides upstream of the transcription start site and a putative translation start codon ATG embedded within a Kozak consensus sequence. Furthermore, based on the deduced amino acid sequence, the extreme amino-terminal coding region contained a classic putative secretory signal sequence . This feature is consistent with the secretory nature of the mucin gene products in the airway and various other organs. Several putative motifs for various transcription factor binding sites were also identified upstream of the transcription start site, as indicated in FIG. 8.

III. *MUC5B* Expression Analysis by Northern Blot

[0099] To further elucidate patterns of *MUC5B* gene regulation, the expression patterns of *MUC5B* in primary and established cultures of TBE-derived human cells were studied using Northern blotting techniques, as described in EXAMPLE 3. *MUC5B* gene expression was analyzed in primary cell lines derived from airway tissues (*i.e.*, TBE cells) as well as in established cell lines, and also in a variety of culture conditions. The established tracheobronchial cell lines used in this study were BEAS-2B, which was derived from SV-40 large T-antigen immortalized bronchial epithelial cells (Ke *et al.*, *Differentiation* 38(1):60-66 [1988]) and HBE1 cells, which are a papilloma virus immortalized tracheal epithelial cell line (Yankaskas *et al.*, *Am. J. Physiol.*, 264:C1219-C1230 [1993]).

[0100] Total RNA was isolated from airway-derived primary cell cultures and established BEAS-2B and HBE1 tracheobronchial cell lines using a guanidinium thiocyanate phenol-chloroform extraction method. A 48-basepair *MUC5B*-specific probe (SEQ ID NO: 3) was derived from the tandem repeat domain of the human *MUC5B* large central exon. The relative abundance of *MUC5B* message in the samples was normalized using an 18S

ribosomal RNA probe. The primary TBE cells were alternatively plated on standard 35 mm tissue culture dishes (TC), collagen-gel coated tissue culture dishes (CG), 25 mm Transwell™ chambers (Corning-COSTAR, Acton, MA; Catalog No. 3506) (BI) or in collagen gel-coated Transwell™ chambers (BICG). The Transwell™ chambers provide a biphasic growth environment where the cells grow in an air-liquid interface that mimics the *in vivo* environment. It is intended that the collagen-gel coating further mimics the *in vivo* environment and provides a more physiological growth environment. These cells were also grown in the presence or absence of retinoic acid.

[0101] As shown in FIG. 4A, primary human TBE cells derived from a "normal" patient expressed detectable levels of *MUC5B* message when cultured in the presence of retinoic acid. The levels of *MUC5B* message in TC and CG cultures were very low compared to the BI and BICG culture conditions, and appeared unaffected by retinoic acid. However, the levels of *MUC5B* message in BI and BICG cultures were greatly enhanced by the presence of retinoic acid, and furthermore, were induced to a level far in excess of the expression observed in the TC and CG culture conditions. This observation is consistent with previous studies (Koo *et al.*, *American Journal of Respiratory Cell and Molecular Biology* 20(1):43-52 [1999] and Wu *et al.*, *European Respiratory Journal* 10(10):2398-2403 [1997]). Thus, *MUC5B* message in culture was affected not only by RA, but also by the culture condition with an order of most-to-least responsive of BICG > BI >>CG >TC. The results of this Northern blot were identical when RNA from cell cultures derived from 11 diseased human tissues were used in place of the TBE cells derived from a normal subject (data not shown). Results on the Northern blot analysis of *MUC*5B message are also consistent with the extent of mucous cell differentiation in these cultures (data not shown).

[0102] Expression of the *MUC5B* gene was also studied in two human TBE immortalized cell lines (HBE1 and BEAS-2B). These cultures were maintained under the BICG culture condition and were maintained in the presence of retinoic acid. Similar to the primary TBE cells, the HBE1 cell line also showed strong *MUC5B* expression, although slightly lower than the TBE culture (*see*, FIG. 4B). For the BEAS-2B subclone S cell line, *MUC5B* expression was undetectable in the Northern blot under all four culture conditions as described above (FIG. 4B, and data not shown).

IV. Mapping of *MUC5B* Transcription Start Site

[0103] A primer extension method was used to map the start site(s) of the *MUC5B* transcription unit, as described in EXAMPLE 5. In this primer extension protocol, total RNA isolated from human trachea tissue or from human primary tracheobronchial epithelial (TBE) cells was reverse-transcribed using a ^{32}P end-labeled primer (the Pel1 primer; SEQ ID NO. 7, and *see* TABLE 2). The radiolabeled reverse-transcribed products were resolved on a denaturing gel simultaneously with a corresponding Sanger (*i.e.*, di-deoxy) sequencing series and DNA size reference markers. The results of the primer extension analysis are shown in FIG. 7. This analysis showed the transcription start site to be located at approximately basepair position 4176, as shown in FIG. 6, and GenBank Accession No. AF107890. Significant degradation and weak signal are observed in this analysis, most likely due to the inherent difficulty in obtaining intact full-length transcripts from genes that have extremely long messages, such as the human *MUC5B* message (Desseyn *et al.*, *Jour. Biol. Chem.*, 273(46):30157-30164 [1998]).

[0104] To overcome the limitations of the primer extension mRNA mapping method of EXAMPLE 5, a modified 5'-rapid amplification of cDNA ends (5'-RACE) method was developed to determine the transcription start site, as described in EXAMPLE 6.

[0105] A 5'-RACE kit (Roche Molecular Biochemicals, Indianapolis, IN) containing a reverse transcriptase was used to synthesize the first-strand cDNA from total RNA (3 μg) isolated from human tracheobronchial tissues or cultures of primary human TBE cells that had been cultured using air-liquid interface culture conditions. Various antisense primers were used to generate first strand cDNA. Instead of 3' tailing with only oligo d(A), the first strand cDNA was also anchored with oligo d(T) by terminal deoxynucleotidyl transferase.

[0106] After tailing, the resulting double stranded cDNA products were used in polymerase chain reactions (PCR) with nested primers within the 3'-end and the 5'-anchor oligo d(T) adapter. PCR amplification was carried out using various primer combinations (*see*, TABLE 2). The resulting PCR products were subcloned into the TA Cloning® vector (Invitrogen, Carlsbad, CA) and sequenced. Since there should be only one common DNA

sequence adjacent to oligo d(T) and oligo d(A) adapters, this DNA sequence should be identical to that of the 5'-end message upstream to the +250/+230 primer. A major advantage of this approach is the use of PCR, which allows the amplification of the 5'-ends of low abundance messages. The sequence analysis of the PCR products generated above identified a transcription start site located located at approximately basepair position 4176, as shown in FIG. 6, and GenBank Accession No. AF107890 (*and see*, FIG. 8). This position is in agreement with the primer extension analysis described in EXAMPLE 5. Both approaches yielded the same conclusion, suggesting that the transcription start site is 18604 basepairs upstream of the large central exon (using the numbering convention of FIG. 8). This putative transcription start site is different from the sites previously reported (Offner *et al.*, *Biochem. Biophys. Res. Comm.*, 251(1):350-355 [1998]; and Van Seuningen *et al.*, *Biochemical Jour.*, 348 Pt 3(12):675-686 [2000]).

V. Construction of *MUC5B* Chimeric Reporter Constructs

[0107] In order to study the transcriptional regulation of the *MUC5B* gene, and also to define minimal promoter elements controlling *MUB5B* tanscription in response to environmental conditions, luciferase reporter constructs under the transcriptional control of *MUC5B* gene sequences were constructed, as described in EXAMPLE 7. The gene sequences used to make these reporter constructs were derived from the isolated genomic DNA described in EXAMPLE 4.

[0108] Fragments of the human *MUC*5B gene corresponding to different 5'-flanking regions as well as a region downstream of the transcription start site (including exon 1) were PCR amplified using appropriate primer pairs (*see*, TABLE 2). The PCR products were subcloned into the promoterless pGL-3 basic vector (Promega, Madison, WI), which contains the luciferase gene open reading frame. Thus, the luciferase gene is under the transcriptional control of the subcloned nucleic acid upstream of the luciferase open reading frame. Three constructs were made, as listed in TABLE 3, and shown in FIG. 9. These reporter constructs, and the *MUC5B* genomic sequences contained in each reporter, were:

[0109] MUC5B-b1 (-1098 to +7). See SEQ ID NO: 31 and FIG. 10.

[0110] MUC5B-b2 (–4169 to +7). See SEQ ID NO: 32 and FIG. 11.

[0111] MUC5B-il (–13 to +2738). See SEQ ID NO: 33 and FIG. 12.

[0112] The MUC5B-b1 and MUC5B-b2 constructs comprise various extents of *MUC5B* sequence upstream of the predicted transcription start site. In addition, the third construct, MUC5B-il, comprises sequences downstream of the presently predicted transcription start site. This last construct was made to test whether these downstream sequences contain elements capable of promoting transcription initiation of the *MUC5B* gene, as proposed in previously published reports (Desseyn *et al.*, *Jour. Biol. Chem.*, 273(46):30157-30164 [1998]; and Van Seuningen *et al.*, *Biochemical Jour.*, 348 Pt 3(12):675-686 [2000]).

[0113] In addition, a MUC5B promoter reporter construct driving the expression of a GFP reporter gene is also provided by the invention. This GFP reporter construct is under the transcriptional control of the –4169 to +7 promoter region (*see*, SEQ ID NO: 32 and FIG. 11). This GFP reporter is analogous to the luciferase reporter MUC5B-b2.

VI. Analysis of *MUC5B* Chimeric Reporter Constructs in Transient Transfection Assays

[0114] The activity of the *MUC5B* reporter constructs described above and in EXAMPLE 7 was assessed in cultured primary TBE cells and established TBE cell lines following transient transfection according to the methods provided in EXAMPLE 8. In addition, the *MUC5B* luciferase reporter activity of the constructs was also assayed in response to various culture conditions. The chimeric reporter plasmids used in the transfections were purified using QIAGEN® plasmid isolation kits, and the transient transfections were done using Roche FuGENE 6™ transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN), all according to the manufacturer's instructions. In these transient transfections, a cotransfected pSV-β-galactosidase (β-gal) expression vector was included for the normalization of transfection efficiency. Cell extracts prepared from the various transfected cell cultures were assayed for both luciferase and β-galactosidase reporter gene activities (*see*, EXAMPLE 8).

[0115] FIG. 13 shows the results of a transfection assay using cultured primary TBE cells and the chimeric *MUC5B* reporter constructs. The primary TBE cells were maintained on standard 35 mm tissue culture dishes (without retinoic acid). As can be seen in the FIG. 13, the reporter gene activity in MUC5B-b1 and MUC5B-b2 transfected cells was two- to five-fold higher, respectively, than those transfected with the promoterless control construct, pGL-3 (labeled "control"). No significant activity was observed in the transfection using the MUC5B-il construct. These results indicate that the regions -1098 to +7 and –4169 to +7 both have promoter activity, and the –4169 to +7 region contains stronger promoter activity than does the -1098 to +7 region. Furthermore, the –13 to +2738 region contained no detectable promoter activity under these conditions.

[0116] FIG. 14 shows an analysis of MUC5B-b2 reporter activity in various cell types, which were primary TBE cells (unfilled bars), HBE1 cells (striped bars) and BEAS-2B (S clone) cells (black bars), all grown in 35 mm tissue culture dishes without retinoic acid. As can be seen in FIG. 14, the MUC5B-b2 promoter was most active in the primary TBE cells, followed by activity observed in the HBE1 cells. No significant promoter activity was observed in the BEAS-2B cells. These results are consistent with the Northern blot data (FIG. 4), which suggests cell type-specific *MUC5B* regulation.

[0117] FIG. 15 shows the results of an experiment examining the effects of cell culture conditions on MUC5B-b2 promoter activity in primary human TBE cells. The TBE cells were maintained in either standard tissue culture dishes (TC) or collagen gel-coated Transwell™ chambers (BICG), and activity of the MUC5B-b2 reporter construct was observed. Furthermore, the cultures were maintained either in the presence or absence of retinoic acid (RA). As can be seen in FIG. 15, when TBE cells were plated on tissue culture dishes, the reporter gene activity was not affected by the addition of retinoic acid. In contrast, the reporter gene activity was elevated five-fold by retinoic acid treatment when transfected cells were maintained under BICG conditions. This culture condition-dependent promoter activity was consistent with the Northern blot data, which showed that culture conditions influenced retinoic acid-dependent *MUC5B* gene expression.

[0118] Thus, the largest of the reporter constructs, MUC5B-b2, contained sufficient *MUC5B* promoter region (*i.e.*, approximately 4 kB) to drive the transcription of the

luciferase open reading frame in a cell type-specific manner. Furthermore, this promoter region was sufficient to respond to various culture conditions, including various growth substratum and nutrient states (*e.g.*, the presence or absence of retinoic acid). These data demonstrate the importance of the biphasic air-liquid interface in regulating *MUC5B* gene expression.

[0119] MUC5B reporter constructs using the GFP open reading frame can also be used to assess promoter activities, both qualitatively and quantitatively. GFP production can be visualized in a fluorescence microscope in either tissues or individual cells as well as quantitated from crude cell extracts prepared from cultured cells or tissues (see, EXAMPLE 10). Furthermore, the expression of luciferase or GFP can also be visualized using immunohistochemical techniques, especially in the analysis of tissue sections.

VII. Construction and Analysis of Transgenic Animals Carrying Chimeric Reporter Constructs

[0120] In order to study the transcriptional regulation of the *MUC5B* gene in the context of a mammalian organism, transgenic animals carrying *MUC5B* reporter constructs were produced using methods well known to one familiar with the art. The reporter constructs used in this study (both luciferase and GFP reporter constructs) are described in EXAMPLE 7. The generation of the respective transgenic mice is described in EXAMPLE 9.

[0121] Transgenic animal technology, including the construction (*i.e.*, establishment) of a desired transgenic animal line (*e.g.*, a mouse line), is common in the art, and the protocols used to establish such transgenic lines are described in many sources (*see, for example*, Hogan *et al.*, *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., [1986]). General discussion of such protocols is provided below. In addition, the actual procedure used to produce the transgenic animals of the invention are provided in EXAMPLE 9. Although the making of transgenic animals is illustrated herein with reference to transgenic mice, this is only for illustrative purpose, and is not to be construed as limiting the scope of the invention. This specific disclosure can be readily

adapted by those skilled in the art to incorporate MUC5B-reporter transgene sequences into any non-human mammal utilizing the methods and materials described herein.

A. Cells Used for Introduction of Transgene

[0122] The transgenic animals of the present invention all include within a plurality of their cells a transgene of the present invention (*e.g.*, a *MUC5B* promoter reporter construct, as described in EXAMPLE 7). In an exemplary embodiment, the transgenic mammals of the invention were produced by introducing a MUC5B-reporter transgene into the germline of the mammal. Embryonal target cells at various developmental stages can be used to introduce a MUC5B-reporter transgene. Different methods are used depending on the stage of development of the embryonal target cell. The specific line(s) of any animal used to practice this invention are selected for general good health, good embryo yields, good pronuclear visibility in the embryo, and good reproductive fitness.

[0123] In one embodiment, the transgene construct is introduced into a single stage embryo. Generally, the female animals are superovulated by hormone treatment, mated and fertilized eggs are recovered. For example, in case of mice, females six weeks of age are induced to superovulate with a 5 IU injection (0.1 ml, i.p.) of pregnant mare serum gonadotropin (PMSG; Sigma) followed 48 hours later by a 5 IU injection (0.1 ml, i.p.) of human chorionic gonadotropin (hCG; Sigma). FVB strain of mice are used in this case. Females are then mated immediately with a stud male overnight. Such females are next examined for copulation plugs. Those that have mated are euthenized by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts and placed in Dulbecco's phosphate buffered saline with 0.5% bovine serum albumin (BSA; Sigma). Surrounding cumulus cells are removed with hyaluronidase (1 mg/ml). Pronuclear embryos are then washed and placed in Earle's balanced salt solution containing 0.5% BSA (EBSS) in a 37.5°C. incubator with a humidified atmosphere at 5% CO_2, 95% air until the time of injection.

[0124] Normally, fertilized embryos are incubated in suitable media until the pronuclei appear. At about this time, the transgene is introduced into the female or male pronucleus as described below. In some species such as mice, the male pronucleus is

preferred. For example, the exogenous genetic material is added to the early male pronucleus, as soon as possible after the formation of the male pronucleus, which is when the male and female pronuclei are well separated and both are located close to the cell membrane. Alternatively, the exogenous genetic material is added to the nucleus of the sperm after it has been induced to undergo decondensation. Sperm containing the exogenous genetic material can then be added to the ovum or the decondensed sperm could be added to the ovum with the transgene constructs being added as soon as possible thereafter.

[0125] In addition to similar biological considerations, physical ones also govern the amount (*e.g.*, volume) of exogenous genetic material, which can be added to the nucleus of the zygote, or to the genetic material which forms a part of the zygote nucleus. Generally, the volume of exogenous genetic material inserted will not exceed about 10 picoliters. The physical effects of addition must not be so great as to physically destroy the viability of the zygote. The biological limit of the number and variety of DNA sequences will vary depending upon the particular zygote and functions of the exogenous genetic material and will be readily apparent to one skilled in the art, because the genetic material, including the exogenous genetic material, of the resulting zygote must be biologically capable of initiating and maintaining the differentiation and development of the zygote into a functional organism.

[0126] The number of copies of the transgene constructs which are added to the zygote is dependent upon the total amount of exogenous genetic material added and will be the amount which enables the genetic transformation to occur. Theoretically only one copy is required; however, generally, numerous copies are utilized, for example, 1,000-20,000 copies of the transgene construct, in order to insure that one copy is functional.

B. Methods of Introducing Transgene

[0127] Each transgene construct to be inserted into the cell must first be in the linear form since the frequency of recombination is higher with linear molecules of DNA as compared to the circular molecules. Therefore, if the construct has been inserted into a vector, linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the transgene sequence.

[0128] Introduction of the transgene into the embryo may be accomplished by any means known in the art so long as it is not destructive to the cell, nuclear membrane or other existing cellular or genetic structures. Some of the widely used methods include microinjection, electroporation, or lipofection. Following introduction of the transgene, the embryo may be incubated *in vitro* for varying amounts of time, or reimplanted into the surrogate host, or both. One common method is to incubate the embryos *in vitro* for about 1-7 days, depending on the species, and then reimplant them into the surrogate host.

[0129] The zygote is the best target for introducing the transgene construct by microinjection method. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2 pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (Brinster *et al., Proc. Natl. Acad. Sci. USA* 82: 4438-4442 [1985]). As a consequence, all cells of the transgenic animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene.

[0130] Retroviral infection can also be used to introduce transgene into a non-human mammal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Jaenich, *Proc. Natl. Acad. Sci. USA* 73: 1260-1264 [1976]). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (*Manipulating the Mouse Embryo*, Hogan (ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, [1986]). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner *et al., Proc. Natl. Acad. Sci. USA* 82: 6927-6931 [1985]; Van der Putten *et al., Proc. Natl. Acad. Sci. USA* 82: 6148-6152 [1985]). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Van der Putten, *supra*; Stewart *et al., EMBO J.*, 6: 383-388 [1987]). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can also be injected into the blastocoele (Jahner *et al., Nature* 298: 623-628 [1982]). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of the

cells which formed the transgenic animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germ line by intrauterine retroviral infection of the midgestation embryo (Jahner *et al.*, (1982) supra).

[0131] Insertion of the transgene construct into the ES cells can be accomplished using a variety of methods well known in the art including for example, electroporation, microinjection, and calcium phosphate treatment. A preferred method of insertion is electroporation, in which the ES cells and the transgene construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer's guidelines for use. After electroporation, the ES cells are typically allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the transgene.

C. Implantation of Embryos

[0132] Pseudopregnant, foster or surrogate mothers are prepared for the purpose of implanting embryos, which have been modified by introducing the transgene. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and it is species dependent. For mice, this stage is about 2-3 days pseudopregnant. Recipient females are mated at the same time as donor females. Although the following description relates to mice, it can be adepted for any other non-human mammal by those skilled in the art. At the time of embryo transfer, the recipient females are anesthetized with an intraperitoneal injection of 0.015 ml of 2.5% avertin per gram of body weight. The oviducts are exposed by a single midline dorsal incision. An incision is then made through the body wall directly over the oviduct. The ovarian bursa is then torn with watchmaker's forceps. Embryos to be transferred are placed in DPBS (Dulbecco's phosphate buffered saline) and in the tip of a transfer pipet (about 10 to 12 embryos). The pipet tip is inserted into the infundibulum and the embryos transferred. After the transfer, the incision is closed by two sutures. The number of embryos implanted into a particular host will vary by species, but will usually be comparable to the number of off spring the species naturally produces.

[0133] Where the ES cell have been used to introduce the transgene, the transformed ES cells are incorporated into the embryo as described earlier, and the embryos may be implanted into the uterus of a pseudopregnant foster mother for gestation.

D. Screening for the Presence or Expression of Transgene

[0134] Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Offspring that are born to the foster mother may be screened initially for mosaic coat color where a coat color selection strategy has been employed. Alternatively, or additionally, screening is often accomplished by Southern blot or PCR of DNA prepared from tail tissue, using a probe that is complementary to at least a portion of the transgene. Western blot analysis or immunohistochemistry using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for screening for the presence of the transgene product. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the RNA expression of the transgene using Northern analysis or RT-PCR.

[0135] Alternative or additional methods for evaluating the presence of the transgene include, without limitation, suitable biochemical assays such as enzyme and/or immunological assays, histological stains for particular marker or enzyme activities, flow cytometric analysis, and the like. Analysis of the blood may also be useful to detect the presence of the transgene product in the blood, as well as to evaluate the effect of the transgene on the levels of various types of blood cells and other blood constituents.

E. Breeding of the Transgenic Animals

[0136] Progeny of the transgenic animals may be obtained by mating the transgenic animal with a suitable partner, or by *in vitro* fertilization of eggs and/or sperm obtained from the transgenic animal. Where mating with a partner is to be performed, the partner may or may not be transgenic; where it is transgenic, it may contain the same or a different transgene, or both. Alternatively, the partner may be a parental line. Where *in vitro* fertilization is used, the fertilized embryo may be implanted into a surrogate host or

incubated in vitro, or both. Using either method, the progeny may be evaluated for the presence of the transgene using methods described above, or other appropriate methods. Typically, crossing and backcrossing is accomplished by mating siblings or a parental strain with an offspring, depending on the goal of each particular step in the breeding process.

F. Cell Lines and Cell Cultures

[0137] The animals of this invention can be used as a source of cells, differentiated or precursor, which can be immortalized in cell culture if desired. Cells containing a MUC5B-reporter can be isolated from the trangenic animal and established *in vitro* as cell lines and used for drug screening. Thus, the transgenic animals of this invention can be used as a source of cells for cell culture. Tissues of transgenic mice are analyzed for the presence and/or expression of the MUC5B-reporter transgene as described, and cells or tissues carrying the reporter transgene are cultured, using standard tissue culture techniques (*see*, EXAMPLE 10).

VIII. Construction and Analysis of Stably Transfected Established TBE Cell Lines Carrying Chimeric *MUC5B* Promoter Reporter Constructs

[0138] The present invention provides a stably transfected established TBE cell line, namely the HBE1 cell line, carrying *MUC5B* reporter constructs (*i.e.*, the constructs described in EXAMPLE 7). Both luciferase and GFP reporter lines were created, where the reporter genes are driven by the MUC5B –4,169 to +7 promoter region. Methods for the construction of the stably transfected cell lines, and a description of MUC5B reporter gene activity in these lines, is provided in EXAMPLE 11. Furthermore, the activity of the stably transfected reporter constructs was analyzed in response to cytokines and environmental stimuli, including interleukin-6 (IL-6), IL-17 and tobacco smoke. It was observed that these stable cell lines expressed detectable levels of the reporter gene, and were strongly induced by the addition of the proinflammatory cytokines IL-6 and IL-17.

IX. Isolation and Analysis of Stably Transfected Primary Cell Cultures Carrying Chimeric *MUC5B* Promoter Reporter Constructs

[0139] The present invention provides compositions and methods for the isolation and reporter gene analysis of stably transfected mouse primary cell cultures carrying the *MUC5B* luciferase or GFP reporter constructs (*i.e.*, the constructs described in EXAMPLE 7). These primary transgenic cell cultures were derived from the transgenic mice described in EXAMPLE 9. This analysis of reporter gene activity included observation of reporter gene activity in response to various culture conditions.

[0140] In one use of these transgenic cells, the transgenic mice were used to isolate TBE cells, which were maintained in culture. The TBE cells were maintained with and without interleukin-6 (IL-6) or IL-17. After a period of time in culture, the cells were harvested, cell extracts were prepared, and luciferase activity was assayed in each cell extract sample. FIG. 16 shows the results of this analysis. As can be seen in the Figure, the addition of the pro-inflammatory cytokines IL-6 or IL-17 to the cell cultures resulted in significant upregulation of the *MUC5B* promoter activity. It is contemplated that this situation mimics the *in vivo* situation, where IL-6 and IL-17 expression are frequently observed in conjunction with infection and other diseases associated with mucin hyperexpression. Thus, it is possible that IL-6 or IL-17 is responsible for the elevated *MUC5B* expression seen in various airway disease states.

X. Compositions and Methods for Cell and Tissue-Restricted Expression of Heterologous Gene Products

[0141] The present invention provides compositions and methods for the cell-type and tissue-restricted expression of a desired gene product. As demonstrated in EXAMPLE 2, *MUC5B* expression is restricted to the epithelia or glandular mucosal surfaces, *e.g.*, the epithelial mucosal surfaces of the airway. It is contemplated that the MUC5B genomic region –4,169 through +7 can direct expression of a cloned downstream gene product to epithelial or glandular mucosal surfaces.

[0142] It is further contemplated that the delivery of certain gene products, other than reporter gene products, under the control of the *MUC5B* –4169/+7 promoter region finds use in the treatment of disease. For example, delivery of a cell-type restricted expression vector encoding an apoptosis-inducing gene product to the cells of a mucinous airway tumor

will suppress and possibly eradicate the tumor in the patient. Furthermore, as expression of the death-inducing gene product can be restricted to glandular mucosal epithelia, the risk of adversely effecting non-glandular mucosal epithelial cells in a patient is minimized.

[0143] In another example, it is contemplated that the -4169/+7 promoter region contains DNA elements that mediate interaction with positive or negative acting transcription factors that control transcription of the *MUC5B* gene (*see*, FIG. 8), and allow the gene to respond to various environmental stimuli, such as growth conditions and the presence of cytokines or other biological agents. Indeed, this is evidenced by the results of experiments described in EXAMPLES 8, 10 and 11. It is contemplated that cell-type specific expression of a negative regulatory protein using a MUC5B-driven expression vector to a patient suffering from a disease characterized by mucus hypersecretion will result in downregulation of mucus production, and therapeutic benefit to the patient. Similarly, expression of an antisense transcript specific for a positive-acting transcription factor (or the MUC5B transcript itself) will also result in therapeutic benefit to a patient suffering from a disease characterized by MUC5B hypersecretion. Antisense technology has been shown to be an effective means for the downregulation of gene expression.

XI. Methods for Drug Screening Using MUC5B Chimeric Reporter Constructs

[0144] The present invention provides novel compositions and methods that find use in the assessment of *MUC5B* gene transcription in response to various culture conditions or treatments. It is contemplated that *MUC5B* reporter constructs can be used to identify compounds which downregulate (*i.e.*, inhibit) *MUC5B* gene expression. Compounds that are able to downregulate *MUC5B* production find use in the treatment of chronic airway diseases characterized by mucin hypersecretion and/or airway plugging. Examples of such diseases include, but are not limited to, cystic fibrosis, bronchial pneumonia, asthma, chronic bronchitis and emphysema. However, it is not intended that the invention be limited to any particular mechanism or mechanisms by which a compound is able to downregulate (*i.e.*, inhibit) *MUC5B* promoter activity. Indeed, it is not necessary to have an understanding of the mechanism or mechanisms controlling *MUC5B* gene regulation in order to make and use the present invention.

[0145] The drug screening methods of the present invention comprise the assessment of activity of a *MUC5B* promoter reporter construct in a suitable cell, in the absence and presence of a test compound. The reporter activities in these two cultures are then compared. A compound that results in the inhibition of the *MUC5B* reporter construct activity is a candidate for further development as a therapeutic agent for the treatment of diseases resulting from mucin, and specifically *MUC5B*, hypersecretion. In a preferred embodiment, the drug screening methods comprise the identification of a compound that is able to inhibit the upregulation of reporter gene activity (*i.e.*, the *MUC5B* hyperexpression) observed in response to various stimuli, such as exposure to IL-6, IL-17 or tobacco smoke. It is contemplated that compounds identified in the screening that are able to inhibit *MUC5B* expression can be delivered to a patient in need of such treatment by oral, parenteral or inhalation means.

[0146] In one embodiment, contacting the compound with the *MUC5B* reporter construct under study will result in at least a 2-fold inhibition of the *MUC5B* promoter activity, preferably at least 5-fold inhibition, more preferably at least 10-fold inhibition, and most preferably at least 50-fold or greater inhibition of *MUC5B* promoter activity.

[0147] The test compound (*i.e.*, a candidate drug) used in the screening is not particularly limited to any type of molecule. However, compounds having low toxicity towards human cells and humans are preferred. A test compound can be organic or inorganic. Test compounds encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate compounds may comprise functional groups necessary for structural interaction with proteins. The candidate compound often comprises cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more functional groups. Candidate compounds are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

[0148] Candidate compounds are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and

biomolecules, including expression of randomized oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs for testing in the methods of the preset invention.

1. Reporter Constructs

[0149] The present invention provides *MUC5B* reporter constructs suitable for use in drug screening protocols. In one preferred embodiment, the present invention provides a luciferase reporter construct driven by *MUC5B* sequences –4169 to +7, relative to the site of transcription initiation (*i.e.*, the MUB5B-b2 reporter construct). This promoter sequence is provided in SEQ ID NO: X, and is shown in FIG. 11. In another embodiment, the present invention provides a luciferase reporter construct driven by MUC5B sequences –1098 to +7, relative to the site of transcription initiation (*i.e.*, the MUB5B-b1 reporter construct). This promoter sequence is provided in SEQ ID NO: X, and is shown in FIG. 10.

[0150] In other embodiments, the present invention provides a green fluorescent protein (GFP) reporter construct driven by *MUC5B* sequences –4169 to +7, relative to the site of transcription initiation. This construct also finds use in drug screening protocols.

[0151] However, it is not intended that the present invention be limited to luciferase or GFP reporter constructs, as the art knows well other suitable reporter genes that find use with the invention. Such alternative reporter systems include, but are not limited to, for example, chloramphenicol acetyltransferase (CAT), β-galactosidase (β-gal), β-glucuronidase (GUS), and secreted alkaline phosphatase (SEAP). Such systems are common in the art, and are described in many sources (*e.g.*, Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology,* Chapter 9, Part II, John Wiley & Sons, Inc., New York [1994]).

2. Cells Finding Use in Methods for Drug Screening

[0152] The present invention teaches the derivation and use of primary cell cultures and established cell lines derived from tracheobronchial epithelial tissue suitable for use in drug screening protocols in conjunction with the *MUC5B* reporter constructs of the invention. In one embodiment, the present invention teaches the isolation and use of primary human TBE cells derived from normal or diseased human subjects (EXAMPLE 1), that find use in drug screening methods of the invention. In another embodiment, the invention teaches the use of primary mouse TBE cells isolated from transgenic mouse lines carrying a *MUC5B* promoter reporter construct (EXAMPLE 10). In another embodiment, the present invention teaches the use of the established HBE-1 cell line (EXAMPLE 8), which also find use in the methods of the present invention. In another embodiment, the invention teaches the use of stably transfected HBE1 cells (EXAMPLE 11).

[0153] However, it is not intended that the present invention be limited to the use of primary TBE cells, or the established HBE1 cell line, as the art knows well numerous other suitable cell cultures and cell lines that also find use with the invention. In fact, it is not intended that the present invention be limited to the use of any particular cell line(s), as many mammalian cell lines also find use with the methods for drug screening of the present invention. The only requirement of such cell lines is that the *MUC5B* reporter constructs of the present invention be active in these cells. Examples of other alternative cell lines falling within the scope of the present invention include, for example but not limited to, the lung-derived lines A549 mucoepidermoid carcinoma cell line, NCI-H292 carcinoma, Calu-3, and Calu-6 (lung carcinoma). Some cell lines from other organs such as HT-29 (colonic cancer) are also common in mucin research, and also find use with the methods of the invention.

3. Cell Culture Conditions Finding Use in Methods for Drug Screening

[0154] The present invention teaches various cell culture conditions suitable for use in drug screening protocols in conjunction with the *MUC5B* reporter constructs. In various embodiments, the present invention teaches cell culture in standard tissue culture dishes (TC), collagen-gel coated tissue culture dishes (CG), Transwell™ chambers (Corning-COSTAR, Acton, MA; Catalog No. 3506) (BI) and collagen gel-coated Transwell™ chambers

(BICG). In a particularly preferred embodiment, the cells are grown in a biphasic, air-liquid interface, as provided in the Transwell™ chambers. In other embodiments, standard tissue culture dishes are used. Furthermore, cultures may be grown in the absence or presence of retinoic acid. Also, cells may be grown in conditions that result in elevated MUC5B gene activity. For example, in some preferred embodiments, the cells are grown in the presence of IL-6 or IL-17 cytokines, or in the presence of tobacco smoke.

[0155] However, it is not intended that the present invention be limited to any particular culture condition(s). The only requirement of the particular culture system is that the culture conditions used result in detectable levels of reporter gene activity expressed from a *MUC5B* reporter gene construct.

4. Cell Transfection Techniques Finding Use in Methods for Drug Screening

[0156] The present invention teaches the use of FuGENE 6™ transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN) in the transfection of cells in the methods of the present invention, all according to the manufacturer's instructions. However, it is not intended that the present invention be limited to the use of FuGENE 6™ transfection reagent, as the art knows well numerous other suitable cell transfection methods that also find use with the invention. Such alternative methods include, but are not limited to, for example, calcium phosphate-DNA co-precipitation, DEAE-dextran mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, recombinant viral infection, biolistics, and proprietary methods sold by various manufacturers. Transfection reagents are available from a large number of manufacturers, including but not limited to, for example, Sigma-Aldrich (St. Louis, MO) and Gibco-BRL-Life Technologies (Gaithersburg, MD). Where viral-based vectors are used, numerous recombinant viral sequences find use with the present invention, including but not limited to adenovirus sequences, adeno-associated virus sequences, retrovirus sequences, herpes virus sequences, vaccinia virus sequences and Moloney virus sequences. Mammalian cell transfection systems are common in the art, and are described in many sources (*e.g.*, Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology,* Chapter 9, Part I, "Transfection of DNA into Eukaryotic Cells," John Wiley & Sons, Inc., New York [1994]).

5. Stable and Transient Cell Transfection Systems Finding Use in Methods for Drug Screening

[0157] The present invention teaches the use of transient and stable transfection of eukaryotic cells using FuGENE 6™ transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN) in the methods of the present invention. In addition, the invention also teaches the use of transgenic animals, as well as cells derived from those animals, that find use in the drug screening methods of the present invention. It is not intended that the present invention be limited to any particular transfection or transgene protocol, as one familiar with the art recognizes that numerous equivalent systems all find use with the present invention. Methods for the transfection of cells and the generation of transgenic animals are common in the art, and can be found described in many sources (*e.g.*, Ausubel *et al.* (eds.), *Current Protocols in Molecular Biology,* Chapter X, Part X, John Wiley & Sons, Inc., New York [1994]).

6. Transgenic Animals Finding Use in Methods for Drug Screening

[0158] The present invention teaches the use of transgenic animals finding use in the drug screening methods of the present invention. The present invention provides transgenic mice carrying MUC5B(-4,169/+7) luciferase or GFP reporter constructs. It is contemplated that such mice can be used directly to assess whether a particular compound has the ability to inhibit *MUC5B* expression.

[0159] In these methods, the reporter gene used in the reporter construct is not particularly limited, but in some embodiments, a luciferase or a GFP gene are used. In one embodiment, the transgenic animal carrying the *MUC5B* reporter construct is a mouse. In this embodiment, the transgenic animal is first treated in such a way as to induce a state of *MUC5B* hyperactivity, and therefor, simulate disease state. For example, it is known that mice treated with certain allergens or tobacco smoke results in a condition characterized by mucin hypersecretion, and thus, provides an animal model for human obstructive airway diseases.

[0160] Once *MUC5B* expression is elevated (or sufficiently detectable), the mouse is administered a candidate compound for testing. The means used to deliver the compound to the animal are not particularly limited, as oral, parenteral and inhalation delivery techniques are all contemplated. In some embodiments, oral administration of the drug is the most preferred method for drug delivery. After a period of time for treatment with the test compound, ranging for example from 1 to 30 days, the mice are sacrificed, and the level of reporter gene activity within that animal's tissues, and in particular, for example, within the airway tissues, is compared in treated versus untreated animals.

[0161] The method of measuring the reporter gene expression in the mouse tissue can be of any suitable method, as taught in EXAMPLE 10. In some embodiments, tissue-sectioning techniques are used. In some embodiments, immunohistochemical analysis is used, where an antibody or a combination of antibodies are used to detect the reporter gene product. In some embodiments, the reporter protein is measured in crude cell or tissue extracts. Compounds that are able to inhibit the expression of the *MUC5B* reporter gene within the transgenic animal are candidates for further development as therapeutic agents for the treatment of diseases characterized by mucin hypersecretion and airway plugging (*e.g.*, in cystic fibrosis or bronchial pneumonia).

[0162] The following EXAMPLES are provided in order to further illustrate certain embodiments and aspects of the present invention. It is not intended that these EXAMPLES should limit the scope of any aspect of the invention.

EXAMPLE 1

Tissue Collection and Cell Culture

[0163] Eleven (11) human tracheobronchial and lung tissue samples were obtained from the University of California, Davis, Medical Center or the Anatomic Gift Foundation (Laurel, MD). All tissue procurement procedures were approved by The Human Subjects Review Committee of the University of California, Davis. Excised tissues were transported to the lab in an ice-cold, minimal essential medium (MEM; Sigma, St. Louis,

MO). A description of the patients from which the samples were taken is shown in TABLE 1, below.

TABLE 1

Patient No.	Age	Sex	Race	Clinic Diagnosis
H311				no lung disease
H313	75	M[1]	C[2]	no lung disease, died of cardiac arrest
H316	45	F	A	no lung disease, died of cardiac arrest
H317	50	M	C	no lung disease
H297				emphysema
H306	66			UIP[3]
H312	62	M		UIP
H314	64			emphysema
H315	57	F	C	UIP
H320	63	M		emphysema
H321	55	F		emphysema

1. M: male, F: female.
2. C: Caucasian, A: African American
3. UIP: usual interstitial pneumonitis

[0164] Tissue samples from the patients listed in TABLE 1 were processed for airway epithelial cell isolation and subsequent culture using techniques known in the art. For example, this procedue is described in Wu *et al.*, *European Respiratory Journal* 10(10):2398-2403 [1997] and Robinson and Wu., *J. Tiss. Cult. Meth.*, 13:95-102 [1991]). Briefly, human surgical or necropsy specimens were obtained and immersed in minimum essential medium (MEM; GIBCO Laboratories) with L-glutamine and without sodium pyruvate or sodium bicarbonate. The specimens were rinsed in this same medium 2 to 5 times, then immersed in a dissociation solution comprising trypsin protease and EDTA overnight at 4°C. The next day, the mucosal surface was washed multiple times with ice-cold MEM with 10% fetal bovine serum. The washes were pooled and centrifuged to isolate the suspended cells.

[0165] The primary tracheobronchial epithelial (TBE) cells contained in the cell pellet were resuspended in a growth medium and cultured in conditions to stimulate a mucoid/ciliary differentiation pathway. This complete serum-free growth medium comprised F-12 or DME/F12 (1:1) media (GIBCO Laboratories) supplemented with insulin (5 μg/ml), transferrin (5 μg/ml), epidermal growth factor (EGF; 10 ng/ml), dexamethasone (DEX; 0.1 μM), cholera toxin (20 ng/ml), bovine hypothalamus extract (BHE; 15 μg/ml), all-trans-

retinoic acid (RA; 30 nM) and calcuim chloride. The medium was changed the following day, and every other day thereafter. The cells were initially innoculated in plastic tissue culture dishes for propagation, and subjected to serial cultivation and passaging as necessary. In general, the primary human TBE cells maintained on plastic culture surfaces were passaged from 1 to 5 times with a total of 20 to 25 population doublings.

[0166] The cultured cells were transferred to various growth substratum and culture conditions, as necessary. In some experiments, the cell suspensions were plated onto standard 35 mm tissue culture dishes (TC), or collagen gel-coated tissue culture dishes (CG). Passage of cells that were plated onto collagen substrate was generally not performed. Some cells are further maintained in a biphasic culture chamber where the cells were maintained in an air-liquid interface. Transwell™ 25 mm chambers (Corning-COSTAR Catalog No. 3506) were used to produce the biphasic culture conditions, although other equivalent systems can also be used, for example, Millipore MILLICELL® culture plates and the Whitcutt culture method (Whitcutt *et al.*, *In Vitro Cell. Dev. Biol.*, 24(5):420-428 [1988]). The biphasic Transwell™ culture chambers can be used without (BI), or with collagen-gel coating (BICG). The use of a biphasic culture system facilitates polarized cell growth, simulating the *in vivo* condition. Furthermore, confluent primary human TBE cells maintained in BICG conditions are known to express mucociliary differentiation markers (Wu *et al.*, *European Respiratory Journal* 10(10):2398-2403 [1997]; Koo *et al.*, *American Journal of Respiratory Cell and Molecular Biology* 20(1):43-52 [1999]; and Bernacki *et al.*, *American Journal of Respiratory Cell and Molecular Biology* 20(4):595-604 [1999]).

[0167] Two immortalized human TBE cell lines were also used in the present studies. These were BEAS-2B subclone S, obtained from Dr. J.F. Lechner (Wayne State University, Detroit, MI), which was derived from SV-40 large T-antigen immortalized bronchial epithelial cells (Ke *et al.*, *Differentiation* 38(1):60-66 [1988]) and HBE1 cells, obtained from Dr. J. Yankaskas (University of North Carolina, Chapel Hill), which are a papilloma virus immortalized tracheal epithelial cell line (Yankaskas *et al.*, *Am. J. Physiol.*, 264:C1219-C1230 [1993]). These cell lines were maintained in serum-free Ham's F12 medium supplemented with six hormonal supplements, which were insulin (5 μg/ml), transferrin (5 μg/ml), epidermal growth factor (10 ng/ml), dexamethasone (0.1 μM), cholera

toxin (20 ng/ml), and bovine hypothalamus extract (15 µg/ml). To induce mucoid/ciliary cell differentiation in these cell lines, retinoic acid (30 nM) was added to the medium, and cultures were maintained in an air-liquid interface, as in the BICG primary culture conditions described above.

EXAMPLE 2

Tissue Fixation and *in situ* Hybridization

[0168] In this example, the tissue samples described in EXAMPLE 1 were fixed, sectioned and probed *in situ* with probes specific for the *MUC5B* and *MUC5AC* transcripts. This example examines the expression of *MUC5B* in mature normal airway tissue, as well as in diseased airway tissue, such as in emphysema.

[0169] *Experimental* - Portions of the tissues described in EXAMPLE 1 were directly fixed in 4% paraformaldehyde at 4°C overnight. The fixed tissues were washed twice using a 50% ethanol solution for 20 min each wash, followed by two additional washes with 70% ethanol. The fixed tissues were then stored in a 70% ethanol solution at 4°C until paraffin block processing. Following paraffin block mounting, the paraffin-embedded tissues were sectioned to a thickness of 5 µm, and mounted to glass slides.

[0170] The fixed and mounted tissue sections were then analyzed by *in situ* hybridization, using techniques known in the art, with antisense oligonucleotide probes corresponding to the tandem repeat units of the human *MUC5B* and *MUC5AC* genes. These probe sequences used were:

[0171] *MUC5B* probe:
5'-TGTGGTCAGCTTTGTGAGGATCCAGGTCGTCCCCGGAGTGGAGGAGGG-3' (SEQ ID NO. 1), and

[0172] *MUC5AC* probe:
5'-AGGGGCAGAAGTTGTGCTCGTTGTGGGAGCAGGGGTTGTGCTGGTTGT-3' (SEQ ID NO. 2).

[0173] These synthetic oligonucleotides (100 pmole each) were end labeled with a digoxigenin oligonucleotide tailing kit (Roche Molecular Biochemicals, Indianapolis, IN), according to the manufacturer's protocol. Sense oligonucleotides corresponding to these

sequences were also synthesized, digoxigenin-tailed and used as a control probe for the hybridization.

[0174] *In situ* hybridization was carried out as per the manufacturer's protocol (Roche Molecular Biochemicals, Indianapolis, IN). Briefly, the glass-mounted tissue sections were digested with 10 μg/ml Proteinase K in 50 mM Tris-Cl, pH 8.0 and 50 mM EDTA for 15 min at 37°C, rinsed twice in 0.2X SSC (where 20X SSC is 3 M NaCl and 0.3 M Na_3citrate, pH 7.0) and then post-fixed in 4% paraformaldehyde/PBS for 20 min. Slides were treated twice for 5 min each wash with 0.1 M triethanolamine, pH 8.0, and blocked by 0.25% acetic anhydride in a 0.1 M triethanolamine (TEA) buffer. The sections were then dehydrated through the ethanol series.

[0175] The fixed glass-mounted tissue sections were then subjected to probe hybridization. Following a prehybridization, a hybridization buffer containing 2X SSC, 1X Denhard's solution, 10% dextran sulfate, 50 mM phosphate buffer (pH 7.0), 50 mM DTT, 250 μg/ml yeast tRNA, 100 μg/ml synthetic polyA DNA (Roche Molecular Biochemicals, Catalog No. 108626), 500 μg/ml salmon sperm DNA, and 0.5 pmol of digoxigenin-tagged oligonucleotide probe (*MUC5B* or *MUC5AC*) was applied to the tissue section slides. The section was hybridized at 45°C overnight in a humidified chamber. Following hybridization, the section was washed twice with 2X SSC for 15 min each wash at 37°C, twice with 1X SSC for 15 min each wash, and twice with 0.25X SSC for 15 min each wash. After the washes, the slide was reacted with anti-digoxigenin primary antibody-alkaline phosphatase conjugate, washed and visualized according to the manufacturer's instructions (Digeoxigenin Nucleic Acid Detection Kit, Roche Molecular Biochemicals, Indianapolis, IN).

[0176] Alcian blue (pH 2.5)-periodic acid-Schiff (AB-PAS) staining, as used in FIG. 2, was done using methods common in the art. The alcian blue acidic reagent was first used to stain acidic mucin proteins as blue. Addition of the periodic acid-Schiff reagent stained neutral mucin proteins as red.

[0177] *Results/Conclusions* - Results of the *in situ* hybridizations and AB-PAS staining are provided in FIGS. 1-3. The panels of FIG. 1 show images of tracheobronchial tissue from a patient with no obvious airway disease or inflammation (Patient No. H316) that have been hybridized with a *MUC5B* probe (SEQ ID NO: 1). The images (FIGS. 1A and 1C)

reveal that *MUC5B* message in a normal subject is mainly expressed on submucosal gland cells of the tracheobronchial tissue. The enlarged picture of the submucosal gland in FIG. 1C supports this conclusion. For surface airway epithelium, *MUC5B* expression was generally very low (FIG. 1A), except in some regions (FIG. 1B). No *MUC5B* message could be demonstrated in the distal airway and parenchyma regions (data not shown). Similar results were also observed in tissue sections from three other patients without diagnosed lung diseases (Nos. H311, H313 and H317).

[0178] In contrast, it was observed that *MUC5B* message was elevated in both the surface epithelium and submucosal glands of tissue sections obtained from a usual interstitial pneumonitis (UIP) patient (No. H312; FIGS. 3A and 3B) and an emphysema patient (No. H297; FIG. 3C), respectively. In FIGS. 3A and 3C, the *MUC5B* message was elevated in both the surface epithelium and the submucosal gland region, in contrast to sections from the "normal" patient (*see*, FIG. 1). Interestingly, *MUC5B* message could also be seen in the surface epithelium of the bronchiole region of the UIP patient (No. H312; FIG. 3B) and emphysema patients (data not shown). Consistently, *in situ* hybridizations using three other emphysema patients and two other UIP patients demonstrated the same results (data not shown).

[0179] FIG. 2 shows airway tissue sections following AB-PAS staining. AB-PAS staining is a pH sensitive staining that differentiates between neutral and acidic mucosubstances (*i.e.*, substances found on or within mucosal surfaces, cells and tissues), including glyco-conjugated proteins. Acidic mucosubstances appear blue following the staining, while neutral polysaccharides stain magenta/red. Thus, goblet cells, which produce mucin proteins and are mucin containing cells, are expected to be AB-PAS positive. In the airways of all the lung disease patients, extensive goblet cell hyperplasia (or metaplasia) in their airway epithelium (FIGS. 2B and 2C) was observed, in contrast to normal airway that had only a few goblet cells (FIG. 2A), based on AB-PAS staining and morphological analysis. The surface expression of *MUC5B* was limited exclusively to the goblet cells, as shown in the FIG. 3.

[0180] These results illustrate the positive correlation between the overexpression of *MUC5B* message by surface epithelial cells and the presence of disease in the airway

region. Such an association was not seen for the expression of *MUC5AC* message (*see*, FIG. 3D). One example of such a comparative study involved seven lung tissue sections from four emphysema and three UIP patients. Representative panels are shown in FIGS. 3C and 3D. In serial tracheal tissue sections from a UIP patient, *MUC5B* message could be seen in both the airway surface epithelium and the submucosal glands (FIG. 3C), while *MUC5AC* message was seen restrictedly in the airway surface epithelium (FIG. 3D) despite an elevated expression. These observations suggest a possible role for MUC5B gene expression in airway goblet cell hyperplasia (or metaplasia), and by extension, in mucin hypersecretion.

[0181] It is known that *MUC5AC* expression is on the epithelial cell surface while MUC5B expression is within the mucus cells of submucosal glands. It is the novel finding of the present invention that *MUC5B* gene expression can be on the epithelial cell surface of patients with chronic airway disease, while in the same patients, the *MUC5AC* gene does not change its expression location even though its expression is also elevated.

EXAMPLE 3

RNA Isolation and Northern Blot Analysis

[0182] To further elucidate poatterns of *MUC5B* gene regulation, the expression patterns of *MUC5B* in primary and established cultures of TBE-derived human cells were studied. This example describes the isolation of RNA and the analysis of *MUC5B* gene expression using Northern blotting techniques. This example analyzes *MUC5B* gene expression in various cultured cell lines derived from airway tissues, and also under various culture conditions.

[0183] *Experimental* - Following the establishment of primary cell cultures from the airway tissues (as described in EXAMPLE 1), the cultures were allowed to expand for 21 days following their plating on the various culture substratum. Total RNA was isolated from the 21-day cultures by a single-step acid guanidinium thiocyanate phenol-chloroform extraction method. Following similar culture conditions, total RNA was also collected from the established BEAS-2B and HBE1 cell lines.

[0184] For Northern blot hybridizations, equal amounts of total RNA (20 μg/lane) were subjected to electrophoresis on a 1.2% agarose gel in the presence of 2.2 mM

formaldehyde, followed by transblotting onto Nytran® nylon membranes (Schleicher & Schuell, Keene, NH) and cross-linked to the membrane using a UV Stratalinker 2400 (Stratagene, La Jolla, CA). The membranes were prehybridized, then hybridized in a solution comprising 6X SSC, 0.5 % SDS, 10 mM EDTA (pH 8.0), 0.5 % disodium pyrophosphate, 5X Denhardt's solution, synthetic polyA DNA (50 μg/ml) and salmon sperm DNA (50 μg/ml). This hybridization included a single-stranded antisense 48 basepair oligonucleotide derived from the human *MUC5B* gene tandem repeat region (*see*, GenBank Accession Number X74955). The probe was end-labeled with γ-^{32}P-ATP by polynucleotide kinase, and had the sequence:

5'-TGTGGTCAGCTCTGTGAGGATCCAGGTCGTCCCCGGAGTGGAGGAGGG-3'

(SEQ ID NO: 3).

[0185] The blots were hybridized overnight (approximately 16 hours) at 55°C. Following hybridization, the blots were subjected to two sets of washes. The first set of washes used a wash solution comprising 2X SSC and 0.1% SDS for two washes for ten minutes each at 55°C. The second set of washes used a wash solution comprising 1X SSC and 0.1% SDS for two washes for 30 minutes each at 55°C. Following the washes, the blots were exposed to either phosphoimaging or autoradiography.

[0186] Following the above analysis for *MUC5B* expression, the blots were stripped, and the relative abundance of *MUC5B* message in the Northern blot lanes was normalized using an oligonucleotide probe specific for the human 18S ribosomal RNA (rRNA) transcript (*see*, GenBank Accession Number X03205).

[0187] *Results/Conclusions* - Northern immunoblots using a *MUC5B* gene probe and various RNA samples, as described above, are shown in FIGS. 4A and 4B. RNA was isolated from primary TBE cells that were alternatively plated on standard 35 mm tissue culture dishes (TC), collagen-gel coated tissue culture dishes (CG), 25 mm Transwell™ chambers (Corning-COSTAR Catalog Number 3506) (BI) or in collagen gel-coated Transwell™ chambers (BICG). The total RNA isolated from these cells was analyzed in the Northern blot, as describe above, and which is shown in FIG. 4A. From FIG. 4A, it can be seen that primary human TBE cells derived from a "normal" patient expressed detectable levels of *MUC5B* message when cultured in the presence of retinoic acid. The levels of

MUC5B message in TC and CG cultures were very low compared to the BI and BICG culture conditions, and appeared unaffected by retinoic acid. However, the levels of *MUC5B* message in BI and BICG cultures were greatly enhanced by the presence of retinoic acid, and furthermore, were induced to a level far in excess of the expression observed in the TC and CG culture conditions. This observation is consistent with previous studies (Koo *et al.*, *American Journal of Respiratory Cell and Molecular Biology* 20(1):43-52 [1999] and Wu *et al.*, *European Respiratory Journal* 10(10):2398-2403 [1997]). Thus, *MUC5B* message in culture was affected not only by RA, but also by the culture condition with an order of most-to-least responsive of BICG > BI >>CG >TC. The results of this Northern blot were identical when RNA from cell cultures derived from 11 diseased human tissues were used in place of the TBE cells derived from a normal subject (data not shown).

[0188] Expression of the *MUC5B* gene was also studied in two commonly used human TBE immortalized cell lines (HBE1 and BEAS-2B). These cultures were maintained under the BICG culture condition and were maintained in the presence of retinoic acid. Similar to the primary TBE cells, the HBE1 cell line also showed strong MUC5B expression, although slightly lower than the TBE culture (*see*, FIG. 4B). For the BEAS-2B subclone S cell line, *MUC5B* expression was undetectable in the Northern blot under all four culture conditions as described above (FIG. 4B, and data not shown).

EXAMPLE 4

Isolation and Characterization of a *MUC5B* Genomic Clone

[0189] This Example describes the isolation of a *MUC5B* genomic clone, and also describes the characterization of the clone, including restriction mapping, sequencing and sequence annotation. The isolated genomic clone comprises 22.7 kB of genomic chromosome 11 sequence. This 22.7 kB sequence includes both *MUC5AC* and 5' *MUC5B* coding sequences, from which it is inferred that the clone must also contain the entirety of the *MUC5B* 5' promoter regulatory region.

[0190] *Isolation of a MUC5B Genomic Clone* - A DNA probe derived from *MUC2* amino-terminal and promoter proximal region sequences was used to screen a genomic cosmid library derived from human placenta (CLONTECH). The probe used in this

screening (SEQ ID NO: 4) corresponded to nucleotide positions 7,081 thru 11,260 of the human *MUC2* genomic sequence provided in GenBank Accession Number U67167. The nucleic acid probe was radiolabeled using Ready-To-Go™ DNA Labeling Beads (Amersham-Pharmacia Biotech, Catalog Number 27-9240-01). The library screening used a bacterial colony lift assay, as widely known in the art, using low stringency hybridization conditions. Bacterial colonies containing library clones were transferred to Nytran® nylon membranes (Schleicher & Schuell, Keene, NH). These membranes were prehybridized, then hybridized with the radiolabelled probe in a solution comprising 6X SSC, 0.5 % SDS, 10 mM EDTA (pH 8.0), 0.5 % disodium pyrophosphate, 5X Denhardt's solution, synthetic polyA DNA (50 μg/ml) and salmon sperm DNA (50 μg/ml). The membranes were hybridized overnight (approximately 16 hours) at 55°C.

[0191] Following hybridization, the blots were subjected to two sets of washes. The first set of washes used a wash solution comprising 2X SSC and 0.1% SDS for two washes for ten minutes each at 55°C. The second set of washes used a wash solution comprising 1X SSC and 0.1% SDS for two washes for 30 minutes each at 55°C. Following the washes, the blots were exposed to either phosphoimaging or autoradiography, and positive clones were identified.

[0192] In view of the amino acid conservation in the 5' end (*i.e.*, amino-terminus) cystine-rich domains between *MUC2* and *MUC5B*, it was contemplated that this approach would identify genomic clones containing the amino-terminal and promoter region of the human *MUC5B* gene. A total of 10^6 cosmid clones were screened, of which eight were positive for hybridization to the *MUC2* probe.

[0193] These eight positive cosmids were subsequently subjected to confirmation in a secondary screen using a Southern blot hybridization with a *MUC5AC* cDNA probe under stringent hybridization conditions. The probe used in this screening step was derived from the 3' end of the MUC5AC gene, and corresponds to nucleotide positions 1,441 through 3,108 of GenBank Accession Number Z48314. The hybridization conditions used in the screening were, specifically,

[0194] 6X SSC, 0.5 % SDS, 10 mM EDTA (pH 8.0), 0.5 % disodium pyrophosphate, 5X Denhardt's solution, synthetic polyA DNA (50 μg/ml) and salmon sperm DNA (50 μg/ml). The blots were hybridized overnight (approximately 16 hours) at 55°C.

[0195] Following hybridization, the blots were subjected to three sets of washes. The first set of washes used a wash solution comprising 2X SSC and 0.1% SDS for two washes for ten minutes each at 65°C. The second set of washes used a wash solution comprising 1X SSC and 0.1% SDS for two washes for 30 minutes each at 65°C. The third set of washes used a wash solution comprising 0.1X SSC and 0.1% SDS for two washes for 30 minutes each at 65°C. Following the washes, the blots were exposed to autoradiography, and positive clones were identified.

[0196] The *MUC5AC* probe was used in this analysis in view of the genetic map of chromosome 11p15.5. That chromosome is suggested to contain a cluster of mucin genes having the order: centromere-*MUC6-MUC2-MUC5AC-MUC5B*. The *MUC2*, *MUC5AC* and *MUC5B* genes all lie on the same strand and are transcribed in the same orientation.

[0197] Thus, a genomic clone containing *MUC5AC* exon sequences, as well as sequences homologous to the *MUC2* promoter-proximal region, may contain sequences from the *MUC5B* promoter region (*see*, Pigny *et al.*, *Genomics* 38(3):340-352 [1996]; Velcich *et al.*, *Jour. Biol. Chem.*, 272(12):7968-7976 [1997]; Meerzaman *et al.*, *Jour. Biol. Chem.*, 269(17):12932-12939 [1994]; and Desseyn *et al.*, *Jour. Biol. Chem.*, 272(6):3168-3178 [1997]).

[0198] Of the eight positive clones identified in the primary screen, only one of those (a single cosmid clone termed Cos-1) was positive in the secondary screening. Sequence analysis of this clone started with the T3 and T7 primer ends of the cosmid backbone to reveal the DNA sequence of both ends of the cloned genomic insert. This sequencing revealed the presence of the 3' end of the *MUC5AC* cDNA and the 5' end of the large central exon of *MUC5B*, respectively. Thus, knowing the gene order 5'-*MUC5AC-MUC5B*-3', the Cos-1 clone should contain genomic DNA that spans the region between the 3' end of *MUC*5AC gene and the 5' end of *MUC*5B coding sequences, and therefor, must also contain the entirety of the *MUC5B* promoter 5' regulatory sequences. The organization of this positive clone is depicted in FIG. 5A. The full length of the genomic DNA insert on

Cos-1 is estimated to be approximately 44 kB, as estimated by restriction mapping. An expanded view of the promoter proximal region and the *MUC5B* exon/intron structure of this region is depicted in FIG. 5B.

[0199] *Restriction Mapping of the MUC5B Cosmid* - Genomic DNA from the Cos-1 cosmid was prepared and digested with *Kpn*I and *Eco*RI restriction enzymes. Southern blotting hybridization was carried out to determine which DNA fragments contain *MUC5AC* gene sequences or *MUC5B* cDNA sequences. The probe corresponding to the 3' end of the *MUC5AC* message is provided in SEQ ID NO: 5 (corresponding to nucleotide positions 1,441 through 3,108 of GenBank Accession Number Z48314). The probe corresponding to the 5' end of *MUC5B* large central exon is provided in SEQ ID NO: XX (corresponding to nucleotide positions 1 through 809 of GenBank Accession Number Z72496).

[0200] DNA fragments that hybridized to the *MUC5B* cDNA probe were isolated and further subcloned by various restriction enzyme digestions into pGem 4Z (Promega, Madison, WI). These subclones were further mapped by restriction enzyme digestion and sequenced. A restriction map of this region is shown in FIG. 9.

[0201] *Genomic DNA Sequencing* - Human genomic DNA in the Cos-1 clone was sequenced using an ABI Prism Model 377 Automated DNA sequencer (Applied Biosystems, Foster City, CA). Various primers corresponding to different regions of the Cos-1 cosmid clone were used in the sequencing. The sequencing data was analyzed and aligned using LaserGene software (DNASTAR, Madison, WI). The genomic sequencing data was used to verify the restriction map and also to establish the exon/intron gene structure. *MUC5B* genomic sequence comprising 22,773 base pairs upstream of the large central exon was generated and submitted to GenBank with the Accession Number AF107890. This 22.7 kB includes all exons/intons upstream of the large central exon, as well as 5' regulatory sequences upstream of the transcription start site. This 22.7 kB sequence is shown in SEQ ID NO. 6, and FIG. 6. This sequence includes 4169 nucleotides upstream of the predicted transcription start site (*see* EXAMPLES 5 and 6, and FIG. 7), as well as 18,604 nucleotides encompassing the 5'-untranslated (5'-UT) region and exon/intron structure from the 5' terminal half of the gene through exon 31 (also termed the large central exon).

[0202] *Sequence Analysis and Annotation* - Among the 22,773 base pairs sequenced, the 5'-most distal 4,169 base pairs correspond to the 5'-flanking region (*i.e.*, the promoter sequence) of *MUC5B*. In addition to the identification of the *MUC5B* transcription start site (*see,* EXAMPLES 5 and 6), other landmarks are also noted in this genomic sequence. Analysis of the sequence revealed the presence of a TATA box 30 nucleotides upstream of the transcription start site and a putative translation start codon ATG embedded within a Kozak consensus sequence. Furthermore, based on the deduced amino acid sequence, the amino terminal peptide contained a classic putative secretory signal sequence (*see*, FIG. 8). This feature is consistent with the secretory nature of the mucin gene products in the airway and various other organs.

[0203] Several putative motifs for various transcription factor binding sites were also identified upstream of the transcription start site, including binding motifs for c-*Myc* at -101, *Ap*-2 at –1,155, *Hoxd*9/10 at –1,189, and *GRE* at. –1,978. In addition, there are two putative motifs for binding of NF-κB (at –237 and –371) and *AP1* (at –497 and –2,000) (*see*, FIG. 8).

EXAMPLE 5

Determination of the *MUC5B* Transcription Start Site by Primer Extension Analysis

[0204] This example describes the identification of the MUC5B transcription start site using a primer extension methodology.

[0205] *Experimental* - A primer extension method was used to map the start site(s) of the *MUC5B* transcription unit. In this primer extension protocol, 50 μg of total RNA was reverse-transcribed using a ^{32}P end-labeled oligonucleotide primer termed Pel1 having the sequence GCGGCACCACGAGCATGGC (SEQ ID NO. 7, and *see* TABLE 2). This primer lies at nucleotide position +123/+105 according to the numbering convention of FIG. 8. The radiolabeled reverse-transcribed products were analyzed on a 6% polyacrylamide gel simultaneously with a corresponding Sanger (*i.e.*, di-deoxy) sequencing series (which used the same Pel1 primer and pcDNA3 vector template) along with DNA size reference markers (pBR322 DNA digested by *Msp*I, New England Biolabs, Inc., Beverly, MA).

[0206] *Results/Conclusions* - Due to the large size of the human *MUC5B* message (Desseyn *et al.*, *Jour. Biol. Chem.*, 273(46):30157-30164 [1998]), the integrity of the isolated *MUC5B* mRNA is difficult to maintain, thus, the primer extension signal is likely to be weak or degraded. The results of the primer extension analysis are shown in FIG. 7. This denaturing PAGE gel contains a Sanger dideoxynucleotide sequencing ladder (in the order GATC) in lanes 3-6 generated using the *fmol*® DNA Cycle Sequencing System (Promega Corporation, Catalog Number Q4100), and also contains radio-labeled DNA size markers indicated on the right. The primer extension reactions are shown in lanes 1 and 2, where lane 1 used RNA template isolated from human trachea tissue, and lane 2 used RNA isolated from human primary tracheobronchial epithelial (TBE) cells. As can be seen in lanes 1 and 2, the primer extension reactions showed the transcription start site to be located approximately at basepair position 4176, as shown in FIG. 6, and GenBank Accession No. AF107890 (*see*, FIG. 8). Significant degradation and weak signal are observed (FIG. 7).

TABLE 2

Method	Primer sequence	Orientation	Position	SEQ ID NO.
5' RACE	GCGGT GCCCA TTGTA CCAGC	antisense	+4106/+4087	8
	TGGAC CAGCG GCAGA CCTCG	nested antisense	+4086/+4067	9
	CAGTC ACCAT GCAGG TCGTAGA	antisense	+1402/+1381	10
	TCATA GGTGG AGATG TGGGC	nested antisense	+1372/+1353	11
	GTGGA AGGGC TTGGG GGTTG ATGAT	antisense	+1997/+1973	12
	GAGAA GGCAC TGTTG GGATC GG	nested antisense	+1960/+1939	13
	TGGGC ATAGA ACTCG TTGAA GG	antisense	+724/+703	14
	GTTGA AGTCC CCACA CAGGC	nested antisense	+692/+673	15
	GGTCT GGTTG GCGTA TTTGG	nested antisense	+668/+649	16
	CTGGG GAAGA CAGTG ACGGG T	antisense	+250/+230	17
	CGGGT GGAAC AAAGC TCACG C	nested antisense	+234/+214	18
	CTGTG GAGCC GAGCT GGGGG A	nested antisense	+162/+142	19
oligo d(T) anchor primer	GACCACGCGTATCGATGTCGACTTTTTT TTTTTTTTTV	sense		20
oligo d(A) anchor primer	GACCACGCGTATCGATGTCGACAAAAA AAAAAAAAAAV	sense		21
RT-PCR	GTGGA AGGGC TTGGG GTTGA TGAT	antisense	+1997/+1974	22
	GAGAA GGCAC TGTTG GGATC GG	nested antisense	+1960/+1939	23
	GGGCC CACAT CTCCA CCTAT	sense	+1351/+1370	24
Primer Extension	GCGGCACCACGAGCATGGC (Pel1 Primer)	antisense	+123/+105	7
Promoter Constructs				
MUC5B-b1	AAGGATCCGGGTGCTTGCTCCCCTGG[1]	antisense (PL1)	+7/-13	25
	AAGCTAGCGCCACGGAGCATTCAGG	sense (PU2)	-1098/-1080	26
MUC5B-b2	AAGGATCCGGGTGCTTGCTCCCCTGG	antisense (PL1)	+7/-13	27
	AAGCTAGCCTGGTTGTGCCTGTCGCTCA	sense (PU1)	-4169/-4149	28
MUC5B-il	AAAGATCTCCAAATTCCAGCCCCTCCAG	antisense (PiL1)	+2738/+2719	29
	AAGCTAGCCAGGGGAGCAAGCACCC	sense (PiU1)	-13/+5	30

Underlined nucleotides are added to the 5'-end of oligonucleotide primers to facilitate cloning. These cloning sites are *Nhe*I (GCTAGC), *Bgl*II (AGATCT), and *Bam*HI (GGATCC), and each is preceded by two "A" residues. V means A or G or C but not T.

EXAMPLE 6

MUC5B Transcription Start Site Mapping Using a Modified 5'-RACE Protocol

[0207] This example describes refined mapping of the start site of the *MUC5B* transcription unit. To overcome the limitations of the primer extension mRNA mapping method of EXAMPLE 5, a modified 5'-rapid amplification of cDNA ends (5'-RACE) method was developed, and is described in the present example.

[0208] *Experimental* - A modified 5'-RACE method was developed to determine the *MUC5B* transcription start site. A 5'-RACE kit (Roche Molecular Biochemicals, Indianapolis, IN) containing a reverse transcriptase was used to synthesize the first-strand cDNA from total RNA (3 μg) isolated from human tracheobronchial tissues or cultures of primary human TBE cells that had been cultured using air-liquid interface culture conditions for at least 21 days. An antisense primer at nucleotide position +250/+230 having the sequence CTGGGGAAGACAGTGACGGGT (SEQ ID NO. 17, and TABLE 2) was used to initiate first-strand cDNA synthesis.

[0209] In the RACE reactions, only a portion of the 5'-most sequence of the transcript is known. Based on that information, a new primer is designed to generate additional PCR products. After tailing, the resulting double stranded cDNA products were used in polymerase chain reactions (PCR) with nested primers within the 3'-end and the 5'-anchor oligo d(T) adapter. These new products are then cloned and sequenced. Still additional primers are designed based on the new sequence, until the 5' terminus of the message is reached. Since every RACE 5' end product is poly-A tailed, if the message start site is A, it will not be detected in the sequencing reactions. To circumvent this problem, the 5' end of the final RACE product was tailed with oligo d(T) by terminal deoxynucleotidyl transferase, instead of 3' tailing with oligo d(A), so that the true start site can be detected. PCR amplification was carried out using the following primers (also see TABLE 2):

[0210] sense oligo d(A) 5' primer:

GACCACGCGTATCGATGTCGACAAAAAAAAAAAAAAAAV (SEQ ID NO. 21)

[0211] sense oligo d(T) 5' primer:

GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTV (SEQ ID NO. 20)

[0212] antisense 3' primer +234/+214:

CGGGTGGAACAAAGCTCACGC (SEQ ID NO. 18)

[0213] antisense 3' primer +162/+142:

CTGTGGAGCCGAGCTGGGGGA (SEQ ID NO. 19)

[0214] The resulting PCR products were subcloned into the TA Cloning® vector (Invitrogen, Carlsbad, CA) and sequenced. Since there should be only one common DNA sequence adjacent to oligo d(T) and oligo d(A) adapters, this DNA sequence should be identical to that of the 5'-end message upstream to the +250/+230 primer. A major advantage of this approach is the use of PCR, which allows the amplification of the 5'-ends of low abundance messages.

[0215] *Results/Conclusions* - The sequence analysis of the PCR products generated above identified a transcription start site located at approximately basepair position 4176, as shown in FIG. 6, and GenBank Accession No. AF107890 (*see*, FIG. 8). This position is in agreement with the primer extension analysis described in EXAMPLE 5, and shown in FIG. 7. Both approaches yielded the same conclusion, suggesting that the transcription start site is 18604 basepairs upstream of the large central exon (using the numbering convention of FIG. 8). This putative transcription start site is different from the sites previously reported (Offner *et al.*, *Biochem. Biophys. Res. Comm.*, 251(1):350-355 [1998]; and Van Seuningen *et al.*, *Biochemical Jour.*, 348 Pt 3(12):675-686 [2000]).

EXAMPLE 7

Construction of Chimeric *MUC5B* Promoter Reporter Constructs

[0216] This example describes the construction of luciferase reporter constructs under the transcriptional control of *MUC5B* gene sequences. Three constructs are described that contain various portions of the *MUC5B* gene promoter region. The gene sequences used to make these reporter constructs were derived from the isolated genomic DNA described in EXAMPLE 4. Assessment of the activity of these constructs is described in EXAMPLE 8.

[0217] Fragments of the human *MUC*5B gene corresponding to different 5'-flanking regions as well as a region downstream of the transcription start site and including exon 1 were PCR amplified using appropriate primer pairs (*see*, TABLE 2 for complete

primer sequences). Total RNA isolated from primary TBE cells grown in an air-liquid interface in a collagen gel in the presence of retinoic acid served as the template for these PCR reactions. The PCR products were digested with appropriate restriction enzymes and subcloned into the promoterless pGL-3 basic vector (Promega, Madison, WI), which contains the luciferase gene open reading frame. Thus, the luciferase gene is under the transcriptional control of the subcloned nucleic acid upstream of the luciferase open reading frame. Clones of these chimeric constructs were verified by DNA sequencing. Three constructs were made, as shown in TABLE 3.

TABLE 3

Construct	Nucleotide Positions	PCR Primer Pairs	Subcloning Sites
MUC5B-b1	-1098 to +7 (SEQ ID NO: 31 and FIG. 10)	PL1 (antisense) SEQ ID NO: 25 PU2 (sense) SEQ ID NO: 26	*Nhe*I/*Bam*HI
MUC5B-b2	–4169 to +7 (SEQ ID NO: 32 and FIG. 11)	PL1 (antisense) SEQ ID NO: 27 PU1 (sense) SEQ ID NO: 28	*Nhe*I/*Bam*HI
MUC5B-il	–13 to +2738 (SEQ ID NO: 33 and FIG. 12)	PiL1 (antisense) SEQ ID NO: 29 PiU1 (sense) SEQ ID NO: 30	*Nhe*I/*Bgl*II

[0218] In addition to the luciferase reporter constructs described above, a *MUC5B* promoter reporter construct encoding a green fluorescent protein (GFP) reporter gene was also constructed. To make this construct, the –4169 to +7 *MUC5B* promoter region was subcloned into a vector backbone (Promega Corporation, Madison, WI) carrying the GFP open reading frame, such that transcription of the open reading frame is under the transcriptional control of the *MUC5B* sequences.

EXAMPLE 8

Transient Transfections and Assessment of Reporter Construct Activity

[0219] This example describes the transient transfection of the *MUC5B* luciferase reporter constructs (*i.e.*, the constructs described in EXAMPLE 7), and the subsequent analysis of their activity in the context of various cell lines and cell culture conditions. This analysis was conducted in primary TBE cells as well as established TBE cell lines, and also in response to various culture conditions.

[0220] *Experimental* - For transient transfection studies, primary TBE cells were cultured in 35 mm dishes and grown to 60-80% confluence. The chimeric reporter plasmids used in the transfections were purified using QIAGEN® plasmid isolation kits, and the transient transfections were done using Roche FuGENE 6™ transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN) according to the manufacturer's instructions. In these transfections, 0.5 μg of MUC5B-luciferase reporter plasmid DNA per 35 mm culture dish was used for each transfection. In addition, 0.5 μg of the pSV-β-gal expression vector was also included in each transfection for the normalization of transfection efficiency between dishes. Following the transfection, cells were cultured for an additional 48 to 72 hours, then harvested.

[0221] Cell extracts were prepared by removing the culture media from the various culture dishes, washing the cells with PBS solution, adding 200 μl of lysis buffer (0.5 M HEPES pH 7.5, 5% Triton-N101, 1 mM $CaCl_2$ and 1 mM $MgCl_2$) directly to each 35 mm dish, incubating and mechanically scraping and removing the contents of the dish. Luciferase reporter gene activity was quantitated using the LucLite™ luciferase reporter assay system (Packard Bioscience/Packard Instrument Company, Meriden, CT) according to the manufacturer's instructions, using a Packard LumiCount™ luminometer (Packard Instruments, Meriden, CT).

[0222] The β-galactosidase reporter gene activity was assayed according to methods known in the art. Briefly, the luciferase cell extracts described above were mixed with an equal volume of β-galactosidase assay buffer (120 mM Na_2HPO_4, 80 mM NaH_2PO_4, 2 mM $MgCl_2$, 100 mM β-mercaptoethanol, 1.33 mg/ml o-nitrophenyl-beta-D-galactopyranoside [ONPG]), then read in a microplate reader (Molecular Devices) at wavelenghth 420 nm.

[0223] For studying the effects of culture conditions on the promoter-reporter gene activity, primary human TBE cultures were grown in 60 mm dishes and transfected with 1 μg of MUC5B promoter-luciferase construct DNA and 0.5 μg pSV-β-gal expression vector. One day following the transfection, cultures were passaged into either 35 mm tissue culture dishes or into collagen gel-coated 25 mm Transwell™ chambers (Corning-COSTAR Catalog Number 3506). Additionally, the cultures were maintained either in the absence or

presence of supplemental all-trans-retinoic acid (30 nM). For Transwell™ cultures, chambers were maintained in an air-liquid interface for an additional three days. Cell extracts were prepared and luciferase and β-galactosidase activities were analyzed as described above.

[0224] For each transfection, relative luciferase activity was expressed after normalization for β-galactosidase activity. The results are presented as a mean of relative activities from at least triplicate dishes, and data is collected from at least three independent experiments. Activity is expressed as units of luciferase activity per unit of β-gal activity (units/beta-gal).

[0225] *Results/Conclusions* - To determine whether the 5' subdomains cloned in EXAMPLE 7 (SEQ ID NOS: 31, 32 and 33, and *see* FIGS. 10-12) contain *cis*-elements sufficient for the initiation or regulation of *MUC5B* transcription, the luciferase reporter constructs were used in transient transfection assays, as described above. The MUC5B-b1 and MUC5B-b2 constructs comprise various extents of *MUC5B* sequence upstream of the predicted transcription start site. These two constructs contain sequences -1098 to +7 (SEQ ID NO: 31) and –4169 to +7 (SEQ ID NO: 32), respectively. In addition, the third construct, MUC5B-il, comprises sequences –13 to +2738 (SEQ ID NO: 33). This construct was made to test whether these downstream sequences contain elements capable of promoting transcription initiation of the *MUC5B* gene, as proposed in previously published reports (Desseyn *et al.*, *Jour. Biol. Chem.*, 273(46):30157-30164 [1998]; and Van Seuningen *et al.*, *Biochemical Jour.*, 348 Pt 3(12):675-686 [2000]).

[0226] FIG. 13 shows the results of a transfection assay using the chimeric reporter constructs shown in FIG. 9 and passage-1 primary TBE cells. The TBE cells were also co-transfected with a β-galactosidase expression vector, and luciferase activity was normalized against β-galactosidase activity to take into account transfection efficiency variability. Relative activities of each of the reporter constructs following transfection in the TBE cells is shown, and activity is expressed as as units of luciferase activity per unit of β-gal activity (units/beta-gal). As can be seen in this FIG. 13, the reporter gene activity in MUC5B-b1 and MUC5B-b2 transfected cells was two- to five-fold higher, respectively, than those transfected with the promoterless control construct, pGL-3 (labeled "control"). However, no significant activity was observed in the transfection using the MUC5B-il

construct. These results indicate that the regions -1098 to +7 and –4169 to +7 both have promoter activity, and the –4169 to +7 region contains stronger promoter activity than does the -1098 to +7 region. Furthermore, the –13 to +2738 region contained no detectable promoter activity under these conditions.

[0227] Based on the above study, the MUC5B-b2 construct was further used to characterize the specificity of the promoter activity. The result of this experiment are shown in FIG. 14. The MUB5B-b2 construct and the pGL3 control construct were transfected into three different cell types, which were passage-1 TBE cells (unfilled bars), HBE1 cells (striped bars) and BEAS-2B (S clone) cells (black bars). As can be seen in FIG. 14, the MUC5B-b2 promoter was most active in the primary TBE cells, followed by activity observed in the HBE1 cells. No significant promoter activity was observed in the BEAS-2B cells. These results are consistent with the Northern blot data (FIG. 4), which suggests cell type-specific gene expression of the *MUC5B* gene.

[0228] In another experiment, as shown in FIG. 15, the effect of cell culture conditions on MUC5B-b2 promoter activity in primary human TBE cells was tested. The TBE cells were maintained in either standard tissue culture dishes (TC) or collagen gel-coated Transwell™ chambers (BICG), and activity of the MUC5B-b2 reporter construct was observed in these cultures. Furthermore, the cultures were maintained either in the presence or absence of retinoic acid (RA). The luciferase reporter gene activity in each transfected culture was normalized to the activity of a cotransfected β-galactosidase expression vector. Results are expressed as "fold increase" of luciferase activity, comparing RA-treated and RA-untreated cultures, where the activity of the RA untreated culture is set to 1. The activity of the MUC5B-b2 reporter in RA-untreated culture in the TC conditions was normalized to 1. Transfections were done in triplicate, and the mean results of two independent experiments are shown.

[0229] As shown in FIG. 15, when transfected cells were plated on tissue culture dishes, the reporter gene activity was not affected by RA. In contrast, the reporter gene activity was elevated five-fold by RA treatment when transfected cells were maintained under BICG conditions. This culture condition-dependent, RA-stimulated promoter activity

was consistent with the Northern blot data, which showed that culture conditions influenced RA-dependent *MUC5B* gene expression.

EXAMPLE 9

Construction of Non-Human Transgenic Animals

[0230] This example describes the construction of transgenic mice carrying luciferase and green fluorescent protein (GFP) reporter constructs driven by the *MUC5B* promoter genomic region –4169 to +7. These constructs are described in EXAMPLE 7. The transgenic mice were made using techniques well known in the art. Briefly, construction followed the following steps:

Egg Production for Injections

[0231] To obtain a large quantity of eggs (>250) for injection, sexually immature FVB/N females (4-5 weeks of age) were superovulated by using consecutive pregnant mare serum gonadotropin (PMS) and human chorionic gonadotropin (HCG) hormone injections. Females were mated to stud males immediately following the HCG injection.

Harvesting Eggs

[0232] Eggs were harvested the next day from the ampulla of the oviduct of the mated females. Eggs were treated with hyaluronidase to remove nurse cells, and were then washed through several dishes of M2 media. Fertilized eggs are then stored in M16 media at 37°C and in 5% CO_2 until injection.

Injection of Eggs

[0233] Approximately 30-50 eggs were removed from the incubator at a time for injection. Under high magnification, each egg is individually injected with a *MUC5B* promoter reporter transgene (either a MUC5B-luciferase reporter or a MUC5B-GFP reporter). After each egg in that group was injected, all eggs were returned to the incubator. This procedure was repeated until all eggs were injected. At the end of the injection period, eggs which did not survive injection were removed from each group.

Implanting the Eggs

[0234] Injected eggs were then implanted in groups of 10-15 bilaterally into the oviduct of pseudopregnant females (females which were mated to vasectomized males). The

animals were allowed to recover from anaesthesia on a warming plate, and then returned to the animal room. Animals were kept under sterile conditions throughout their pregnancy, and the implanted mothers were brought to term.

Selection of Transgenic Progeny

[0235] Progeny of the implanted mothers were analyzed for the presence of transgene sequences using a combination of PCR and Southern blotting techniques with tail DNA. Mice demonstrating germ line transmission of transgene sequences were identified. The transgenic mice were maintained as heterozygotes. Multiple lines of mice that stably inherit MUC5B-luciferase and MUC5B-GFP transgene sequences were identified and independently maintained.

EXAMPLE 10

Analysis of *MUC5B* Reporter Constructs in Transgenic Animals

[0236] This example describes the analysis of *MUC5B* promoter reporter constructs carried as integrated transgenes in mice. The construction of these mice is described in EXAMPLE 9. The expession of these reporter genes is analyzed using two different protocols (*i.e.*, one for luciferase activity analysis, and one for GFP analysis). Furthermore, the activity of these reporters is studied in response to various cytokines and environmental factors, such as interleukin-6 (IL-6), IL-17 and tobacco smoke.

A. Analysis of Reporter Gene Activity in Primary TBE Cultures Derived from Transgenic Mice

[0237] The transgenic mice described in EXAMPLE 9 were used to isolate TBE cells, which were maintained in culture. The TBE cells were maintained in three culture conditions, which were control (no supplement), with interleukin-6 (IL-6) at a concentration of 10 ng/ml or with IL-17 at a concentration of 10 ng/ml. The cells were maintained in the presence of the cytokines for 7 days, harvested and cell extracts were prepared as described in EXAMPLE 8. The luciferase activity in each cell extract was determined, and normalized for total protein concentration of the extract samples.

[0238] FIG. 16 shows the results of this analysis of the MUC5B-b2 luciferase reporter activity. As can be seen in the Figure, the addition of the pro-inflammatory

cytokines IL-6 or IL-17 to the cell cultures resulted in significant upregulation of the *MUC5B* promoter activity. It is contemplated that this situation mimics the in vivo situation, where IL-6 and IL-17 expression are frequently observed in conjunction with infection and other diseases associated with mucin hyperexpression. Thus, it is possible that IL-6 or IL-17 is responsible for the elevated *MUC5B* expression seen in various airway disease states.

B. Analysis of Reporter Gene Activity in Tissues Derived from Transgenic Mice

[0239] Alternatively, and in a manner similar to that described above, reporter gene activity can be analyzed in cultured cells isolated from any particular tissue from the transgenic animal. For example, it is contemplated that cultured colon tissue epithelial cells can also be used in a manner as described in this EXAMPLE, as colon tissue has been demonstrated to produce mucin proteins *in vivo*, and is also a suitable system for the study of *MUC5B* gene regulation.

[0240] In another alternative protocol, analysis of reporter gene activity in cells of a particular tissue isolated from the transgenic animal can be done directly by generating protein extracts from tissues isolated from the transgenic animals. Samples of these tissue extracts can be analyzed for the presence of reporter gene, for example, using the same luciferase assay as described in EXAMPLE 8. In a related protocol, the presence of GFP can also be quantitated in a crude protein extract using a suitable scintillation fluid (*e.g.*, FloroCount, Packard Bioscience) and a fluorescence excitation detection apparatus.

C. Analysis of Reporter Gene Activity in Tissue Sections Derived from GFP-Reporter Transgenic Mice

[0241] In another alternative protocol, GFP reporter gene activity in the cells of any particular tissue isolated from a transgenic animal carrying a MUC5B-GFP reporter construct can be assessed by fluorescence microscopy. For example, tissues can be isolated from a transgenic mouse carrying the MUC5B-GFP reporter construct, and this tissue is sectioned and mounted to glass slides. These sections are then observed under a suitable excitation fluorescence microscope, and the GFP protein can be visualized.

D. Analysis of Reporter Gene Activity in Tissue Sections Derived from Transgenic Mice Using Immunohistochemistry

[0242] In another alternative protocol, reporter gene activity in the cells of any particular tissue isolated from a transgenic animal carrying a MUC5B promoter reporter construct can be analyzed by immunohistochemistry using a primary antibody to the particular reporter gene product encoded by the transgene. For example, anti-GFP and anti-luciferase antibodies are commercially available (*see, e.g.*, Goat Anti-Luciferase Polyclonal Antibody, Promega Corporation, Catalog No. G7451). The bound primary antibody can then be detected using a suitable secondary antibody (*e.g.*, Donkey Anti-Goat IgG Alkaline Phosphatase Conjugate, Promega Corporation, Catalog No. V1151), and thus, expression of the reporter gene in the tissue sections can be visualized.

EXAMPLE 11

Construction and Analysis of Stably Transfected Established Cell Lines Carrying *MUC5B* Promoter Reporter Constructs

[0243] This example describes the stable transfection of the –4,169/+7 MUC5B-luciferase and MUC5B-GFP reporter constructs (*i.e.*, the constructs described in EXAMPLE 7) into the established TBE cell line HBE1.

[0244] *Experimental* – The established cell line HBE1 was cultured in 35 mm dishes and grown to 60-80% confluence. These cells were cotransfected with either MUC-5B reporter construct and a second plasmid encoding the neomycin-resistance (*neo*) selectable marker. The chimeric reporter plasmids used in the transfections were purified using QIAGEN® plasmid isolation kits, and the cotransfections were done using Roche FuGENE 6™ transfection reagent (Roche Molecular Biochemicals, Indianapolis, IN) according to the manufacturer's instructions. In these transfections, 2.5 μg of MUC5B reporter plasmid DNA and 0.5 μg of the neomycin resistance marker plasmid per 35 mm culture dish were used for each transfection.

[0245] Following the cotransfection, cells were cultured for an additional 48 to 72 hours. At this time, the medium was replaced with fresh medium containing the neomycin analogue G-418 at a concentration of 100 μg/ml. The selection was maintained for approximately 21 days, at which time clones of resistant transfected cells were replated and maintained as continuous lines. Cell extracts were prepared and luciferase activity

quantitated exactly as described in EXAMPLE 8, with the exception that cell extracts were normalized for total protein content, and not β-galactosidase activity. In addition, these cells were cultured in the absence or presence of IL-6 (10 ng/ml) or IL-17 (10 ng/ml). It was observed that these cells expressed detectable luciferase activity, and this activity is upregulated when cells are cultured in the presence of IL-6 or IL-17.

****

[0246] All of the references identified herein, including patents, patent applications, and publications, are hereby incorporated by reference in their entireties.

[0247] While the invention has been described with an emphasis upon preferred embodiments, it will be obvious to those of ordinary skill in the art that variations in the preferred method, compound, and composition can be used and that it is intended that the invention can be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the following claims.